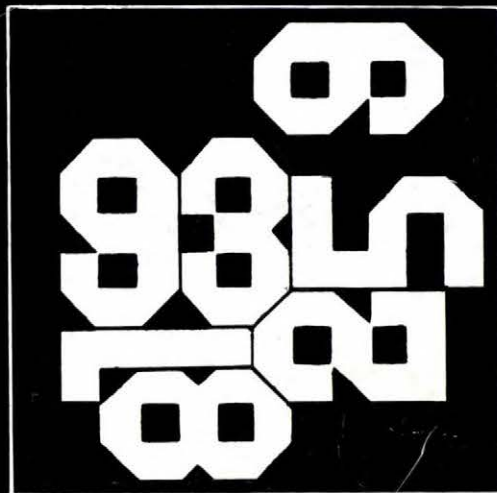


MTA Számítástechnikai és Automatizálási Kutató Intézet Budapest





MAGYAR TUDOMÁNYOS AKADÉMIA  
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

**A magyar nyelv elemzése  
számítógéppel**

(Tervek egy természetes nyelvű interfészhez)

Bach Iván  
Farkas Ernő  
Naszódi Mátyás

Tanulmányok 199/1987

A kiadásért felelős:

*Dr. KEVICZKY LÁSZLÓ*

Főosztályvezető:

*CSABA LÁSZLÓ*

*A tanulmány az OMF 911-13-1 szakfeladat 61-42-044/86  
szerződése alapján készült.*

ISBN 963 311 230 3

ISSN 0324-2951

MTA SzTAKI 1987. aug. 1.

SZÁMALK

## Tartalom

<b>Bevezetés</b> .....	5
A természetes nyelv feldolgozásának alapproblémája .....	5
A magyar nyelv szabályai .....	7
A tervezett program felépítése .....	8
Az eddig elért eredmények .....	10
<b>A magyar szavak morfológiai elemzése 8</b> .....	12
A szó .....	12
Szóelemzés .....	12
Morfológiai szeletelő .....	13
Elvi modell .....	13
Gyakorlati modell .....	14
Kivitelezés .....	20
Tapasztalat .....	20
<b>A szótár</b> .....	21
Morfológiai jellemzők .....	21
Szintaktikai jellemzők .....	22
Szemantikai jellemzők .....	23
A magyar szófajok morfológiai - szintaktikai felosztása ....	24
<b>A szintaxis alapjai</b> .....	27
Igei szerkezetek .....	28
Formális definíció .....	30
Névszói kifejezés .....	34
Formális leírás .....	35
Egyeztetés .....	38
Az alany, a tárgy és az ige .....	38
A birtokos szerkezet egyeztetése .....	41
Vonzatok .....	43
Argumentum és vonzat .....	44
Kötelező és opcionális argumentumok, defaultok .....	46
A szintaxis analízis megvalósítása .....	47
<b>Hivatkozások</b> .....	50
<b>Szemantikus reprezentáció és szemantikus jegyek</b> .....	53

## Mellékletek

<b>Melléklet a kísérleti morfológiai elemző toldalékairól</b> .....	55
<b>Kísérleti szótár</b> .....	59

**A magyar nyelv elemzése számítógéppel**  
(Tervek egy természetes nyelvű interfészhez)

## Bevezetés

Az előző "A magyar nyelv alkalmazása a számítástechnikában" című előtanulmányunkban áttekintést adtunk arról, hogy elsősorban angol és orosz minták alapján, hogyan képzelhetjük el egy nyelvészeti alapokon nyugvó természetes nyelvi interfész felépítését.

Mi korábban programnyelvek fordító programjaival, programnyelvi és természetes nyelvű szövegszerkesztő programokkal foglalkoztunk. Ezen tapasztalatok birtokában állíthatjuk, hogy egy természetes nyelvű interfész program létrehozása nem látszik technikailag különösebben nehéz feladatnak. Az áttekintett programok kapcsán sehol nem találkoztunk számunkra teljesen új és meglepő programtechnikai megoldásokkal. Annál nehezebbnek látszik a feladat tartalmi oldala, a magyar nyelv szabályainak lefordítása olyan formális szabályokká, amelyeket azután gépre lehet vinni.

### A természetes nyelv feldolgozásának alapproblémája

Amikor Noam Chomsky felállította a nyelvek leírására szolgáló első 1957-es modelljét egy olyan formalizmust állított fel amely alapján a jelsorozatok egyértelműen két csoportba sorolhatók: vannak olyanok, amelyek mondatok és így a nyelvhez tartoznak, és vannak olyanok, amik nem. Mi, akik hosszú éveket dolgoztunk programozási nyelveket fordító programokon, tudjuk, ez a felosztás, így megvalósíthatatlan. Még a programozási nyelvekben is, amelyek pedig gondosan megtervezett emberi alkotások, vannak olyan konstrukciók, amelyeken a szakemberek vitáznak, hogy vajon legálisak és értelmesek-e. Hatványozottan így van ez a természetes nyelvek esetén. A mondatok, szavak, szóalakok, kifejezések döntő többségéről megállapítható ugyan, hogy helyes-e vagy helytelen, de igen széles sáv marad még közöttük, amelyről vitatkozni lehet.

Azt hihetnénk esetleg, hogy ez a bizonytalanság talán nyelvi képzetlenségünkől fakad, és a szakemberek számára kevés az ilyen probléma. Az esetek többségében ennek szinte az ellenkezője igaz. Mint ismeretes a nyelvészeti munkákban a helytelen alakot az előtte álló \*-gal jelölik meg, a vitathatót pedig, a szerző indulatától függően, egy vagy több kérdőjellel. Tapasztalatunk szerint minden nyelvészeti munkában jónéhány kérdőjel akad. Ennek a bizonytalanságnak több forrása van, az egyik a különböző környezet, neveltetés, stb., a másik a nyelv állandó változása.

A nyelv ilyen laza felfogásával élesen ellentétben áll a számítógép alkalmazása, ha azt mondatok analizálására vagy szintézisére akarjuk felhasználni. Ebben az esetben maga a program élesen eldönti, hogy melyik mondatot tekinti helyesnek és értelmezhetőnek, és melyiket hibásnak. A számítógép számára vitatható mondat nincs.



A fentiekből következik, hogy a számítógépes program nem tudja és nem is akarja leírni a nyelvet olyan mélységben (fejlődésében és ellentmondásaiban), ahogy azt a nyelvészet teszi, hanem egy modellt készítünk, amelyről eleve tudjuk, hogy a magyar nyelvet csak részben fedi le. Már most biztosak vagyunk benne, bizonyos nyelvi konstrukciókat (mint pl. a mondat- átszövés, régies, de még használt igealakok, stb.) nem fogunk kezelni.

Nyelvi modellünk nemcsak a teljes nyelv átfogása tekintetében lesz korlátozott, hanem feltehetően meglehetősen durva is lesz, és így a helyes mondatok közé valószínűleg becsúsznak olyan mondatok is, amelyek noha hibásak, a program mégis jó mondatként értelmezi őket. A programozási nyelvek körében ortogonalitásnak nevezzük azt a tulajdonságot, ha egy szabály kivételek és más szabályokra való tekintet nélkül alkalmazható. Ortogonális szabályrendszerre nyilvánvalóan könnyebb programot írni. A természetes nyelv viszont (mint azt már a morfológiáról szóló fejezetben is láthatjuk) egyáltalán nem ortogonális. A fontosabb kivételeket és a szabályok egymásrahatását természetesen kezelni kell, de az összes kivétel és egymásrahatás kiderítése és a programba való beépítése megvalósíthatatlannak látszik.

Felmerülhet a kérdés, hogy egyáltalán megoldható-e a feladat, hiszen ha a modellt túl bőre szabjuk, sok nem korrekt mondat is bele fog kerülni, ha viszont túl szűk sok helyes magyar mondat kimarad belőle. Erre azt a választ adhatjuk, hogy mi a gyakorlat számára akarunk programot készíteni, és ebből a szempontból a számítógépes kommunikáció két alfeladatra bomlik: a bemenő szöveg analízise és megértése, és a válaszok generálása. Ha abból indulunk ki, hogy a természetes nyelvű interfészt azok akarják használni, akiknek ez a nyelv a természetes, akkor feltételezhetjük, hogy a felhasználók ritkán fognak nyelvtani hibákat elkövetni a géphez forduláskor, viszont feltehetőleg nem tetszene nekik, ha a gép rossz magyarsággal válaszolna. Így hát világos, hogy két különböző modellt kell felállítanunk a két feladatra, egy tág lehetőségeket adót a bemenetre, és egy szűkebb választékút, de nyelvtanilag pontosabbat a válaszokra.

Mi jelen munkánkban csak az analízissel foglalkozunk, és feltételezzük, hogy a bemenő szöveg nyelvtanilag alapjában véve helyes. Ezért bizonyos vizsgálatokat a modellünkből teljesen kihagytunk. Tapasztalatunk szerint teljesen fölösleges például a szavakat magánhangzó-illeszkedés szempontjából vizsgálni, noha ennek szabályai és a kivételek ismertek. Ilyen hiba a beírt szövegekben praktikusán nem fordul elő, ezért a ráfordított emberi és gépi munka teljesen felesleges lenne.

Általában is azt az elvet követjük, hogy a nyelvtani szabályokból csak annyit veszünk figyelembe, amennyi a megértéshez szükséges, és elhanyagoljuk azokat, amelyek az analízis számára nem adnak újabb információt.



Lehet persze, hogy néhány ponton feltételezésünk csal, és a későbbiek folyamán bizonyos nyelvi jelenségek vizsgálatára mégis rákényszerülünk, amelyeket korábban feleslegesnek hittünk. Igyekszünk modellünket úgy megcsinálni, hogy a későbbiekben könnyen módosítható és fejleszthető legyen. Ahol lehet arra törekszünk, hogy a nyelv modell és a hozzákapcsolódó program ne folyjon össze. Ha a táblázatokban leírt nyelvmodellt módosítjuk, szűkítjük vagy bővítjük, akkor a programon ne kelljen változtatni, vagy csak újabb jóldefiniált interfészű függvényeket kelljen hozzávenni.

### A magyar nyelv szabályai

Ha lett volna a magyar nyelvnek olyan viszonylag teljes formális leírása, amelyből kiindulhattunk volna, akkor a létrehozandó eszközöket bizonyára ehhez igazítottuk volna. Így azonban, azt az utat jártuk, hogy megpróbáltuk összegyűjteni a magyar nyelv szabályait, és megkíséreltük végiggondolni, hogyan lehet az összefüggéseket a gépben ábrázolni, és a szabályokat géppel ellenőrizni, végül megkezdjük a szabályok formalizálását. Erre olyan formalizmust vezettünk be, amely alkalmasnak látszik a magyar nyelv géptől és programtól független absztrakt leírására, és ugyanakkor ebből a formális leírásból generálható az a program, amely elvégzi a természetes nyelvű szöveg analízisét.

A leírás alapjául a következő forrásokat használtuk:

Bencédy József - Fábián Pál - Rácz Endre - Velcsov Mártonné:  
A mai magyar nyelv, Hatodik kiadás 1985.

A magyar helyesírás szabályai, Tizenegyedik kiadás 1986.

É. Kiss Katalin:

A magyar mondat szerkezet generatív leírása, 1983.  
Configurationality in Hungarian, 1987.

Hadrovics László:

A funkcionális magyar mondat tan alapjai, 1969.

Prószéky Gábor, Kálmán László, Kornai András és társaik számos cikke

Antal László számos műve

Kiefer Ferenc:

Az előfeltevések elmélete, 1983.

Pléh Csaba számos munkája az anafórikus kapcsolatokról

Felhasználtuk továbbá:

a nagy Magyar Értelmező Szótárat, a Gyakorisági Szótárat

Papp Ferenc: Szóvégmутató szótár

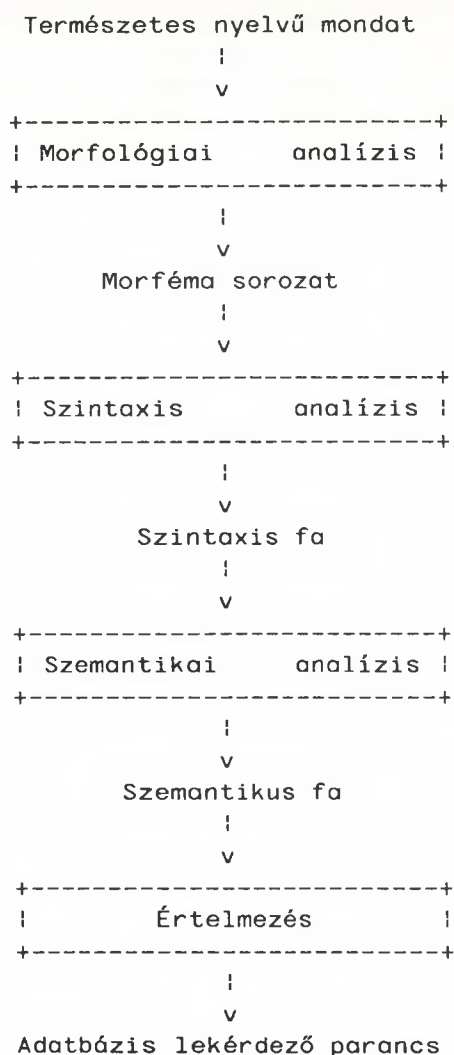
Аpreszjan Ju.D. - Páll Erna:

Orosz ige - magyar ige Vonzatok és kapcsolódások, 1982.

Továbbá köszönetet kell mondanunk mindazon nyelvészeknek, akik segítségünkre voltak: elsősorban Füredi Mihálynak, Prószéky Gábornak, valamint Varga Dénesnek.

### A tervezett program felépítése

Munkánkban a nyelv különböző síkjain különböző mértékben haladtunk előre. Hogy erről részletesebben beszélhessünk, vissza kell idéznünk az előtervében lefektetett terv diagramját, amely programunk vázlatos felépítését mutatataja:



A morfológiai analízis a bemenő szöveg szóalakjait morfémák (szótövek és végződések) sorozatára bontja. Az egyes morfémákhoz egy-egy kódszámot rendel hozzá, amely megmutatja az adott morféma grammatikai szerepét. (Pl. ige, főnév, melléknév, stb. illetve kijelentő módú, jelenidejű egyesszám 3. személyű igerag, határozói igenév képzője, névszói többesszám jel, melléknévi fokozás, stb.) A felbontás során természetesen foglalkozni kell a teljes hasonulással és más hasonló jelenségekkel is.

A szintaxis analízis egy mondat morféma sorozatát egy fa alakú gráffá alakítja. A fa levelein a szótövek illetve a végződések állnak. A nyelvtanilag összetartozó elemek egy csomópontban kapcsolódnak össze, ezek még nagyobb nyelvtani egységeket (szintagmákat) alkotnak, és azok még nagyobbakat, amíg végül is létre jön a mondat. A szintaktikus fa a mondat mondattani szerkezetét írja le.

A következő fázis ezt a fát egy olyan fává alakítja át, ami már a mondatban szereplő tartalmi kapcsolatokat írja le. Az átalakítás során az egyik feladat, hogy fel kell oldanunk olyan hivatkozásokat (például a névmásokat), amelyekkel egyszerűsített formában hivatkozunk bizonyos korábbi dolgokra. A másik feladat az, hogy bizonyos egyszerűsítéseket, standardizálásokat hajtsunk végre a mondat ábrázolásán. Például, véleményünk szerint, ha azt mondjuk, hogy "neki el lehet mennie" vagy azt mondjuk, hogy "elmehet" a két kijelentés azonos, noha külső formáját tekintve meglehetősen sok eltérést vehetünk észre. Kézenfekvő, hogy a szemantika szintjén a két kijelentést azonosan ábrázoljuk. Ezenkívül, ellenőrizni kell, hogy a mondat nem csak nyelvtanilag helyes, hanem a benne leírtak végre is hajthatók. Ezeket a tevékenységeket együtt nevezzük a továbbiakban szemantikus analízisnek.

(A szemantikáról több féle értelemben szokás beszélni. Nem vonjuk kétségbe, hogy valamilyen szinten lehet meg lehet adni a teljes magyar nyelv szemantikus leírását, de ez a leírás igen általános lenne, és utána egy külön lépésben kellene kiválasztani a helyes magyar mondatok közül azokat a mondatokat, amelyek az adott feladatnál értelmesen felhasználhatók. Mi eleve csak azokat a mondatokat tekintjük helyesnek, amelyek az adott helyzetben használhatók, és egy, az adott témához és feladathoz kötött, részletes szemantikus ábrázolásra gondolunk.)

A program utolsó lépése a mondat szemantikus reprezentációját adatbázist lekérdező parancsokká fordítja.

### Az eddig elért eredmények

Mint azt fentebb említettük, a különböző területeken a munkában különböző mértékben jutottunk előre. Ennek több oka is van. Az első ok, hogy a munka fázisai, mint azt az ábra is mutatja szekvenciálisan egymásra épülnek, és amíg a megelőző fázis nem állít elő elfogadható eredményt, addig a következő fázis kipróbálása nem kezdődhet meg.

Ez fokozottan így van esetünkben, azért is, mert a magyar nyelvre vonatkozó számítógépes tapasztalatok hiányában nem tudjuk jól megítélni az egyes problémák súlyát.

A másik ok az, hogy még mindig nem döntöttük el, mi lesz az a konkrét adatbázis, amelyre az interfészt készítjük. Így mindazoknál a pontoknál ahol a szemantika szóba jön, csak igen általános elképzeléseink vannak, mivel terveink szerint a szemantikus jegyeket a téma és a várható feladatok függvényében akarjuk megválasztani.

Eredményeinket a következőkben foglalhatjuk össze:

A morfológiai analízis megvalósítására elkészült két kísérleti program, amellyel sok kísérletet hajtottunk végre, és meglehetősen jó eredményeket értünk el. Ezen tapasztalatok alapján készült el, a most már véglegesnek szánt morfológiai elemző rendszerterve, és valószínűleg ez év szeptemberére készen lesz a végleges program is. Ez a program a korábbiaknál több végződést (azon belül képzőt) lesz képes kezelni (így a szótár mérete csökkenhet), és képes lesz kezelni egy sor olyan jelenséget, mint a teljes hasonulás, a tű-rövidülés, a harmadik azonos mássalhangzó kiesése, stb (így a szótárban csökkenhet az egy szóhoz tartozó tőalakok száma).

Elkészült egy kísérleti szótár, amely körülbelül 2500 tőalakot tartalmaz és ezek a később ismertetett elvek alapján szófajokra vannak osztva. Ez a szótár szükséges a morfológiai elemzéshez, és az itt kapott szófaji kódokból indul ki a szintaktikai elemzés. A szótárnak kb. 700 eleme esik egybe a gyakorisági szótár leggyakoribb szavaival.

Elkészült a magyar nyelv szintaxisának formális leírása kissé módosított attribútum-nyelvtan segítségével. A nyelvész kollégák ezt a leírást elolvasták, és elfogadhatónak tartják. Ennek ellenére biztosak vagyunk benne, hogy a leírás tele van hiányosságokkal, de hogy mik ezek, és melyiknek milyen súlya van, majd csak szeptember táján az első próbafuttatások során állapíthatjuk meg. A próbához meg kell írni az elemző program három szükséges alkotórészét: a kötött szórendű szabályok kezelését, a szabad szórendű szabályok kezelését és a vonzatok kezelését, valamint össze kell állítani a vonzatszótárat. Eddig csupán a kötött szórendű szabályok ellenőrzésére történtek kísérletek.

Pontos (bár nem formálisan leírt) terveink vannak az egyeztetések elvégzésével kapcsolatban.

A vonzatok és a hivatkozásokkal kapcsolatos terveink általános jellegűek. Ezek pontosabb leírásához szükség lenne ugyanis a szemantikus jegyek halmazának definiálására.

A következő fejezetekben összefoglaljuk a felmerült problémákat, az ezekkel kapcsolatos döntéseket és terveket. Hangsúlyozni kívánjuk, hogy a magyar nyelv itt megadott leírása csak egy a számos lehetséges felfogás közül, és a gyakorlat fogja megmutatni, hogy mennyire használható. Előbb azonban ki kell küszöbölni a benne található hibákat és hiányosságokat.



## A magyar szavak morfológiai elemzése

A következőkben a feladat meghatározása után egy elvi (alapvetően generatív) modellt vázolunk fel, majd az általunk kidolgozott gyakorlati analitikus modellt ismertetjük. Végül néhány tapasztalati tanulságot vonunk le.

### A szó:

Szónak (a nyelvészek szóalaknak hívják) nevezzük az írott szöveg egybeírt (csak betűből és kötőjelből álló, más írásjelet, szóközt nem tartalmazó) darabját, mely vagy maga egy szótő, vagy egy szótőből a későbbiekben leírandó szabályokkal létrehozott egység. (A számjeggyel és más írásjellel jelzett egységeket nem tekintjük szónak.) A magyar nyelv ún. agglutináló nyelv, ami azt jelenti, hogy a szótőhöz végződéses tapadhatnak, és néhány előtag is kapcsolódhat a szóhoz. Ezenkívül a szóösszetétel is lényeges szóalkotási mód. A szavak különböző szempontokból osztályokba sorolhatók (lásd később: szófajok, ragozási típusok). A toldalékok (előtagok) szemantikai módosításon kívül a szavakat egyik szóosztályból a másikba viszik. Különböző szavaknak lehet azonos alakjuk. A szavak szintaktikai és szemantikai tulajdonságaival későbbi fejezetben foglalkozunk.

### Szóelemzés:

A szóelemzés célja, hogy megállapítsa azt, milyen szótőből és milyen toldalékolással (és előtagokkal), esetleg milyen szóösszetétellel jött létre az adott szó (pl. *legelemibb: leg-elem-i-bb*). Egyes nyelvekben, mint például az angolban, ahol a toldalékolás és előtagok alkalmazása nem létezik vagy szegényes, a szóelemzés gyakorlatilag egy szótár alkalmas kiépítésével megoldható. Nem így a magyarban, ahol a toldalékolás szó is kaphat toldalékot. (Ilyen esetekben relatív szótőről beszélhetünk.) A toldalékok szekvenciája eléggé kötött. A toldalékok (előtagok) is osztályokba sorolhatók aszerint, hogy milyen szóosztályokra alkalmazhatók és milyen szóosztályba kerül az új szó. A különböző toldalékoknak lehet azonos az alakjuk (pl. *kutyá-nak lát-nak*). Egyes toldalékok módosíthatják a relatív szótövet (tőváltozatok, hasonulások). Ennek alapján, négy feladatot kell megoldani:

- A toldalékok (előtagok) levágása.
- A toldalékolás következtében végbement tőváltozások "visszacsinálása".
- A toldalékok és szóosztályok egyeztetésének ellenőrzése.
- A szótövek és a toldalékok és azok osztályának meghatározása.

Az első hárommal a morfológiai szétválasztás foglalkozik, míg a negyedik alapvetően szótár segítségével történik. Bár a négy funkció szétválasztható, a megoldás során összefonódnak, egymást kiegészítik.

### Morfológiai szeletelő:

A morfológiai elemző modul a természetesnyelvi elemző rendszer szerves része. Célja, az írott szavak morféimákra bontása a további (szintaktikai, szemantikai) elemzést megelőzően. Mivel a teljes nyelvi rendszer interaktív ember-gép kapcsolat részére készül, a válaszüidők rövidsége érdekében lényeges az elemző nagy sebessége. A rendszer az általunk köznyelvinek tartott alakokkal foglalkozik, és nem foglalkozik régies, vagy tájnyelvi alakokkal, de alkalmas bővítésére, szűkítésre módosítására is, ha ilyenre lenne szükség. A kidolgozott módszer alkalmas morfológiai szabályok leírása alapján gyors elemző előállítására.

### Elvi modell:

A morfológia a következő hat fogalommal operál: ABC, Szóosztályok, Tőváltozások, Tőtár, Előtagtár, Toldaléktár.

**ABC:** a nyelvre jellemző. (Az ABC rendezettsége számunkra nem lényeges.)

**Szóosztályok:** a szavak csoportosítása szófaj, hangrend, ragozástípus alapján.

**Tőtár:** tartalmazza az összes használható szó tövét, jelölve, mely szóosztályba tartozik.

**Toldaléktár:** tartalmazza az összes használható toldaléket jelölve, hogy milyen osztályú szóra (szófajra, hangrendre stb.) alkalmazható, milyen osztályú szófajt állít elő és milyen jellegű tőváltozást idézhet elő.

**Előtagtár:** tartalmazza az összes használható előtagot jelölve, hogy milyen osztályú szóra (szófajra, hangrendre stb.) alkalmazható, és milyen osztályú szófajt állít elő. (A magyar nyelv előtagjai nem okoznak tőváltozást.)

**Tőváltozások:** bizonyos szótövek bizonyos toldalékok esetén megváltoznak. Ilyen változástípusok a magánhangzó kiesések, betoldások, rövidülések, nyúlások, hasonulások (ide sorolható az is, amikor nem a tő, hanem a toldalék módosul) stb. Ezek a változások magyar nyelvben mindig a (relatív) szótő utolsó egy-két betűjét érinti.

**Toldalékolás:** egy adott (relatív) szótőre alkalmazható egy toldalék, ha a (relatív) szótő szóosztálya megegyezik a toldalék alkalmazhatósági szóosztályával. Ekkor először végrehajtjuk a tövön a toldalékhoz tartozó tőváltozást majd az így kapott módosított tő után fűzzük a toldaléket, és eredményül a toldalék által előírt szóosztályba tartozó szót kapunk.

**Az előtag alkalmazása:** a fentihez hasonlóan definiálható.

**Elemzés:** Az elemzés feladata, hogy egy adott szóról megállapítsa, milyen szótőből, milyen toldalékolási (előtag) szekvencia, esetleg milyen szóösszetétel segítségével jöhet létre. Könnyen belátható, hogy ha az üres toldalék nem okoz rekurzivitást, valamint a szótőváltozások a toldalékkal mindig növelik a szó hosszát, márpedig ez a magyar nyelvre igaz, akkor az elemzés egy véges automatával véges lépésben elvégezhető.



### Gyakorlati modell:

A fenti megállapítás nem jelenti azt, hogy az elemzés minden részét a gyakorlatban is végesautomata-modell alapján kell kivitelezni. Ha az összes szóbagyővő szótővet is véges automatával szeretnénk felismerni, minimális szótár esetén is egy többtízezer állapotú véges automatára lenne szükség, amelynek mind a létrehozása, mind a karbantartása teljesíthetetlen feladat volna.

Hasonlóan rossz megoldás lenne, ha az elemző mindent keresgetéssel, próbálgatással kezelne (ilyen elemzők léteznek, és elsősorban pontos nyelvhelyességi illetve demonstratív céllal készültek), hisz a hatékonyság nem lenne kielégítő.

Az elemzés történhetne a szótőtől a toldalékok egyenkénti levágásán keresztül balról jobbra, de mások tapasztalatával megegyezően úgy véljük, hogy ez elég gyakori szótárhoz való fordulással és több hibás részelemzéshez vezetne.

Problémát okozhat a tövváltozások kezelése is. Az Apreszjan és csoportja által létrehozott orosz elemzőben nincs algoritmikus tövváltozáskezelés, hanem a szavak tövváltozataival együtt szerepelnek a szótárban. Ez a módszer a magyar nyelv esetén azért sem követhető, mert toldalékolt tö is szenvedhet tövváltozást.

Legmegfelelőbbnek az a megoldás tűnik, hogy a toldalékolási szabályokkal ellentétes irányban, a szóalakok végéről egy megfelelő véges automata leválaszt egy végződést, amely egy (párszáz elemű) toldalékosztály egyik lehetséges eleme, majd végrehajtja az esetleges tövváltozás inverzét; ha előtag is lehet, azt leválasztja a szó bal oldaláról, végül a lehetséges tövet illetve töveket a több ezer elemű szótárból keresi ki a morfológiai-lexikai elemző. Ily módon a szótári keresés minimálisra csökkenthető.

Lényeges módosítások vezethetők be az algoritmus gyorsítása érdekében annak tudatában, hogy elemzőt készítünk, és feltételezhetjük, hogy nyelvtanilag helyes szavakat kell vizsgálni. További egyszerűsítésekre is találunk lehetőséget a magyar nyelv esetében (pl. az előtagok nem okoznak tövváltozást).

A szavak és toldalékok egyeztetésénél mi alapvetően csak a szófaji egyeztetést vesszük figyelembe, és néhány a szó végét érintő hangtani egyeztetést (pl. a mult idő milyen betű után milyen alakú lehet). A hangrendi és egyéb ragozási szabályok egyébként sem egészen egyértelműek (*mászm-mászok, mondta-mondotta, tőszavak-tőszók, adapter-ok, adapter-ek ...*).

Sok helyen érdemes volt eltérni a nyelvészeti szokásos nyelvtantól és létrehozni egy pragmatikus szóképzési modellt, annak tudatában, hogy a morfológia csupán egy előfázisa a "megértésnek", és mi egy nyelvhelyességi elemző programot kívánunk írni. Ily módon a modell a következőképpen módosul:

**Szóosztályok:** Az osztályozás alapvetően szófaji, illetve toldalékolási tulajdonságok szerint történik.

A szótári elemzés szempontjából három fő osztályba sorolhatók a szavak: ige, névszó, egyéb.

A morfológiai elemzés során további alosztályokat használunk:

Az ige lehet ragozott és ragozatlan, igekötős és igekötő nélküli (a ragozott igék közé soroljuk a főnévi és határozói igeneveket is, mivel ezek után már nem lehet újabb toldalék).

A névszók felosztása aszerint történik, hogy milyen típusú toldalékot kapott (tehát milyen toldalék szelhető le), így névszói tő, különböző mértékben fokozott, többesszámú, illetve birtokjellel ellátott, birtokosjellel ellátott és esetraggal ellátott.

Az egyéb kategóriába taroznak a nem toldalékolható tövek (pl. névelő, kötőszó).

Ezek részletesebb felhasználását lásd a toldalékoklás szekvenciáinál.

Ettől függetlenül, a szavak ragozási és hangtani szóosztályokba sorolhatók, Ez utóbbiak közül legismertebb a hangrendi osztályozás, mely alapvetően meghatározó a illeszkedő toldalék kiválasztásánál, de más nehezebben elemezhető hangtani illeszkedések is szerepet játszanak. A hangrendi és ragozástípus szerinti felosztás elemzőnkben nem játszik szerepet, helyette elegendőnek bizonyult a szóalakok végét vizsgálni. (Lásd toldaléktár.)

**Szótár:** nemcsak szótöveket tartalmaz. Érdemes külön elemként felvenni az olyan szóösszetételeket, képzett szavakat is, melyek értelme nem vezethető le a szóalkotás módjából. A szótóhossz így sem haladja meg a 25 karaktert. Tartalmaz ezen kívül olyan ragozott alakokat is, melyek teljesen rendhagyó módon képződnek (pl. *az*, *annak*), és ezt a speciális ragozást nem érdemes bekódolni a toldaléktárba. Mindezen bővítések miatt a szótár várhatóan 30 - 50 százalékkal nő. A szóalakon kívül a szótár tartalmazza a szó faját, ragozástípusát. Praktikus okokból a szótár tartalmaz egy olyan információt is, hogy érdemes-e tovább darabolni az adott szóalakot. Részletesebb leírást lásd a Szótár című fejezetben.

**Toldalék és előtagtár:** tartalmazza az összes levágható toldalékot aszerint osztályozva, hogy hova tartozik az adott toldalékkal ellátott szó. Az esetleges kötőhangokat a toldalékok részeinek tekintjük. A toldaléktárban az is jelölve van, hogy milyen töv változást idézhet elő, és milyen típusú szó marad a toldalék levágása után. Ilyen értelemben a következő toldaléktípusok vannak:

**Igeragok:** ragozott igeről vágható le, és levágása után ragozatlan ige marad. Ez tulajdonképpen egy összevont toldalék, amely magába foglalja az időt, módot, személyt és számot (pl. *mond-ott-át-ok* helyett egyszerűen *mond-ottátok*). Ide soroljuk a főnévi igenévi képzőket, és a határozói igeneveket is (*mond-anom*, *mond-va*), mivel ezeket sem követhet további toldalék.

**Ige-ige képzők:** ragozatlan igeről választható le, és a maradék is ragozatlan ige; csak a *-hat*, *-het*, *-gat*, *-get*, *-tat*, *-tet*, *-at*, *-et*.

**Ige-névszó képző:** névszói töről választható le, és a maradék ragozatlan ige (pl. *dob-ás*).

**Esetragok:** ragozott névszóról választható le, és a maradék ragozatlan névszó (pl. *tojás-t, ló-ként*).

**Birtokosjel:** ragozatlan névszóról választható le, és a maradék birtokosjel nélküli névszó (pl. *fül-é*).

**Birtokjel-többesszámjel:** egy kategóriába vettük, mivel egymást kizáróan ugyanolyan pozícióban szerepelhetnek. Birtokosjel nélküli névszóról választható le, és a maradék fokozott tő (pl. *ház-ak, piros-aim*).

**Középfok jele:** fokozott töről választható le, és a maradék névszói tő (pl. *kék-ebb*). Vele együtt vágandó le a *leg-*, *legesleg* előtag.

**Névszó-névszó képző:** névszói töről választható le, és a maradék névszói tő (pl. *nyak-ú, kék-es*).

**Névszó-igei képző:** igtöről választható le, és a maradék fokozott tő (pl. *kék-ít*).

**Igekötő:** Előtag, mely ragozatlan igeről választható le, és a maradék igező (pl. *le-húz*).

A toldaléktárban az is jelölve van a toldalékoknál, hogy a relatív szótő (az esetleges töváltozás után) milyen betűre végződhet (pl. az igei múlt egy *t*-je csak bizonyos mássalhangzó után állhat). Ez redundáns információ a ragozási csoportok megjelölése mellett, de gyakorlatilag elégséges az elemzéshez a ragozási csoportok megjelölése nélkül is. Tapasztalatunk szerint elemzésnél a hangrendi osztályozás sem játszik szerepet. A gyakorlatban rendkívül ritkán fordul elő olyan szóalak, mely végéről formálisan úgy vágható le toldalék, hogy az így kapott, nem valódi relatív vagy abszolút szótő hangrendileg ne illeszkednék a toldalékhoz (mint például *balek bal-ek*). Ráadásul éppen az egyik leggyakoribb vita és félreírás abból származik, hogy a szavak hangrendje nem jól meghatározott (pl. *adapter-ok vagy adapterek*). Lényeges információ az is, hogy egy adott toldalékosztály előtt milyen szótőváltozás lehetséges. A toldaléktár teljesen gyakorlati céllal készült, ezért nem feltétlen követi a hagyományos magyar nyelvi toldalékolást. Külön toldaléknak tekintjük a kötőhangos és a kötőhang nélküli toldalékokat (*-t, -ot, -et, -öt*). Nem szerepelnek benne a régies, ritkán használt toldalékok. Kihagytuk az olyan toldalékokat is melyek szemantikailag nem elemezhetőek (*-ság, -ség, -odalom, -edelem* stb.), viszont szerepelhetnek benne mesterséges ragok, melyeket az egyszerűbb kezelés érdekében vezetünk be (pl. a *lás-sál* szóban a *-sál* külön igerag). Bizonyos toldalékösszevonásokat is alkalmazunk hasonló célból (pl. a *-nom* főnévi igenév egyszám első személyű alakja szétválaszthatatlan toldalék).

**Tőváltozások:** A töváltozások helyett jobb azok inverzeit kódolni, tehát azt, hogy pl. *irodalm* alakból *irodalom* szót kell visszatranszformálni bizonyos toldaléklevágások esetén. A korábbiak alapján nem kell minden töváltozást felvenni, de azokat feltétlen, melyek a toldalékolt relatív szótövek esetén is előfordulnak. Ha csak az általunk kezelendő toldalékokat tekintem, akkor ilyen (relatív) szótőváltozás három típusú van a magyarban: a szóvegi magánhangzónyúlás (*fűzet-e, fűzet-é-ből*),



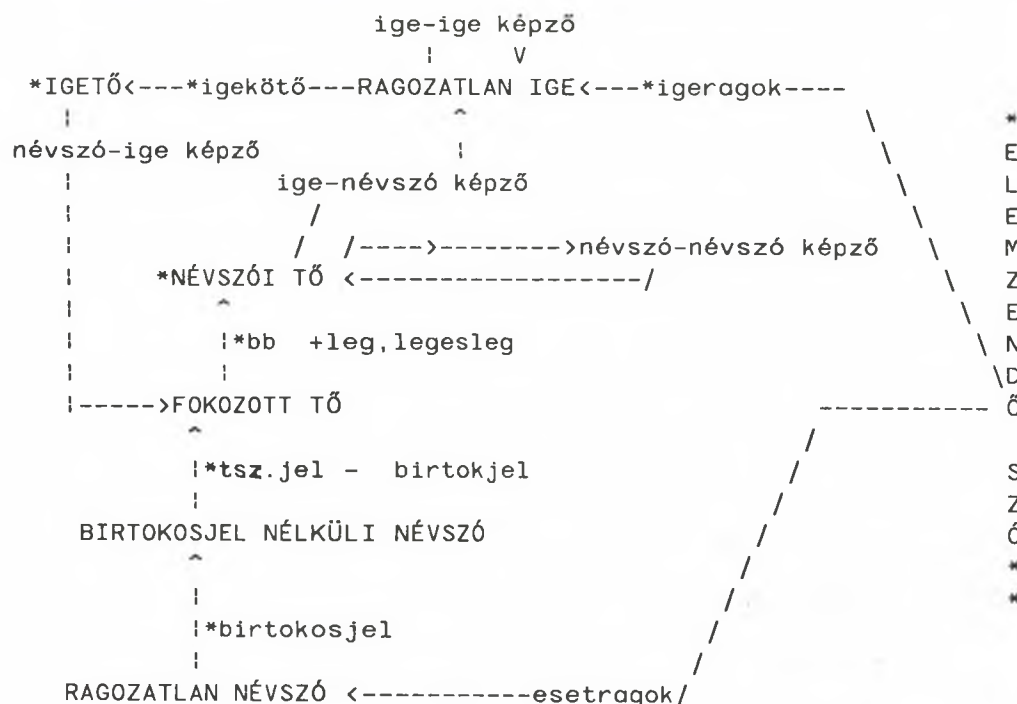


A csomópontok (NAGYBETŰVEL) a szóosztályoknak, az élek (kisbetűvel) a toldalékosztályoknak (előtagosztályoknak) felel meg. A \*-gal jelölt toldalékok lehetnek "üres" toldalékok is. Bár a felírt gráf alapján a toldalékok száma nem korlátozott, a gyakorlatban 3-4 toldaléknál többet nem használnak a magyar nyelvben, sőt bármiféle rekurzió mesterkélt (káposztás-ít-ott-talan-ít-ottátok), de az elemzés során nem zárható ki a ciklus. P1: kékihíthatatlanul

elem alak	kategória	szóosztály	szóalak
kék	szótő	névszói tő	kék
---	fokozó	fokozott tő	kék
-ít	névszó-ige képző	igető	kékit
---	igekötő	ragozatlan ige	kékit
-het	ige-ige képző	ragozatlan ige	kékihíthat
-etlen	ige-névszó képző	névszói tő	kékihíthatatlan
---	fokozó	fokozott tő	kékihíthatatlan
---	számjel	birtokosjel nélküli tő	kékihíthatatlan
---	birtokosjel	ragozatlan névszó	kékihíthatatlan
-ül	esetrag	ragozott névszó	kékihíthatatlanul

Ezzel a szekvenciamodellel jól elemezhetőek a magyar szavak, de elvileg hibás alakok, illetve hibás elemzések is keletkezhetnek. Pontosabban, mivel szófaji felosztásunk rendkívül durva, létezhetnek olyan hibás szóalakok, melyeket elemzőnk értelmez, pedig nincsen értelme. Ez azt jelenti, hogy a valós nyelvnél jóval bővebbet fogad el az elemző. A gyakorlatban ennek kicsi a valószínűsége.

Az inverz (elemző) gráfból is jól látszik, hogy nem fordul elő ürestoldalék ciklus.



Ez a gráf az elemzés vázát adja meg, pontosabban megadja, hogy az egyes toldalékosztályoknak megfelelően automaták hogyan vannak felfűzve. Az elemzés az ELEMZENDŐ SZÓ csomópontnál indul el. A \*-gal jelölt csomópontoknál kell ellenőrizni, hogy a szó szerepel-e a szótárban. Így háromféle szót kereshet a szótár rutin:

**Teljes alakú** (esetleges toldalék levágása nélküli) **szó**: lehet bármilyen okból a szótárba felvett ragozott vagy ragozatlan alakú névszó vagy ige, de lehet az egyéb szófajú szóalak is.

**Névszói tő**: a gyakorlatban ez lehet ragozott tő is.

**Igető**: ez is lehet ragozatlan, és a rendhagyó ragozás miatt ragozott ige.

A gráf alapján elemezhető a magyarban használt formák döntő többsége. Az ily módon nem elemezhető formák olyan ritkán használtak (mint pl. *al-hat-nék-om-ban*), hogy feltételezhetjük, hogy az alkalmazási témakörben nem lesz rájuk szükség. Ha mégis, a gráf egyszerű bővítésével építhetők be ezek a formák a morfológiai elemzőbe. A gráf alapján természetesen elemezhetők olyan mesterségesen kitalált szavak is, melyek a magyar nyelvben elő sem fordulhatnak (pl. *fék-ebb-ít*), de a gyakorlatban ez a struktúra helyes szavaknál nem vezet alternatív rossz felbontáshoz. A helyes megoldás minden esetben előáll.

**Elemzés**: Az adott szó végéről levágja a lehetséges végződéseket a toldalékolás sorrendi szabályainak megfelelően, az elejéről az előtagokat, miközben figyelembe veszi a lehetséges töváltozások inverzét is. Az így kapott szótövet esetleg felbontja szóösszetételként, és az így kapott szótöveket azonosítja a szótárral. Szóosztályegyeztetés esetén eredményül a szótövet, az előtagot és a toldalékokat adja vissza. Egy szó több megoldást is adhat, egyrészt mert a tő és a végződés is lehet többértelmű, másrészt mert a szó többféleképpen vágható szét töre és toldalékokra. Az elemző minden lehetséges megoldást felsorol.

A toldaléktár, illetve előtagtár felhasználásával egy programgenerátor az előfázisban toldalékosztályonként egy-egy véges automatát generál, és ezeket az automatákat összefűzi a szóosztályok közötti átmeneteknek megfelelően. (Így lényegében egy BTN jellegű rendszert kapunk.) Ily módon a toldalékok halmaza és osztályozása az elemzőbe "be van drótozva" míg a szótár adat jellegű információ. Az elemző bemenete egy szó, kimenete egy lista melynek elemei a lehetséges szótórkódból, és toldalékolási szekvenciából állnak. A szótórkód tartalmazza a szótárban talált szó kódját, mely a szintaktikai-szemantikai elemzéshez szükséges (lásd szótár). A toldalékolási szekvencia az alkalmazott toldalékok kódját tartalmazza.

#### Tapasztalat:

A kezdeti próbálkozások alapján hatékony és rugalmasan generálható módszert dolgoztunk ki. Felmerül a kérdés, mennyire fedí a valós szóelemzést algoritmusunk. A mind szótani, mind hangtani illeszkedési szempontból elnagyolt szabályok miatt a helyes szóalakok helyes elemzését mindenképpen megtalálja az elemző. A kérdés inkább az, hogy keletkeznek-e félreelemzések, vagy nem. Mivel feltételezzük, hogy csak helyes szavakat kell elemezni, a többértelműség problémáját kell vizsgálni. A látszólagos felületes szóosztályozás (lényegében csak névszói és igei csoportok kezelése) és a ragozási osztályok "semibevétele" ellenére pontos morfológiai felosztásokat kapunk. Az elemző elvileg helytelenül is elemezhet. A többértelműségnek több oka van:

1. Maga a szótó többértelmű: *lép* mint a méh viaszháza, és mint ige. Ezen a problémán általában nem jelentkezik, ha a szó toldalékolat, de a példánkban szereplő szónál ez sem segít, hiszen a *-nek* lehet igerag is, és esetrag is. A szövegben a leggyakoribb többértelműség az *az* határozott, és az *egy* határozatlan névelő formailag megegyezik az *az* utalószóval, illetve az *egy* számnévvel.

2. A toldalék többértelmű: *-ják* mint kijelentőmódú igerag és mint felszólító módú igerag.

3. Opcionális töváltozás következménye: *gyümölcs-é-t*, *gyümölcs-e-t*, mármint a gyümölcs ízét, vagy Péter gyümölcsét.

4. Képzett alak sajátos jelentéssel: *leves*, *lé-s=leves* (*leves gyümölcs*, *gyümölcsleves*).

5. Tévesen alkalmazott töváltozás, pl: a *szén-szenet* mintájára *kenet=kén-et*, mint a *kén* tárgyesete.

6. A felületes modell miatti (hangrendi illeszkedés semmibe vétele, egyszerűsített szóosztályozás) téves levágás, pl: *balek=bal-ek* a *bal* többesszáma.

A kísérleti próbálkozások alapján (2000-3000-es szótár, 70%-os toldalékkészlet) az elemző által többértelműen felbontott szavak túlnyomó többsége valóban többértelmű (az első négy csoportba tartozik). A százaléokban ki nem fejezhető tévedések (utolsó két csoport) várhatóan jól tisztázhatóak a szintaktikai-szemantikai elemzésnél.



## A szótár

Az adott munkához szükséges szavakat egy szótárban gyűjtjük össze. Pillanatnyilag az a feltevésünk, hogy ha a téma megfelelően korlátos, akkor 2000-3000 szó elegendő a dialógushoz.

Kérdés persze, hogy mit értünk szó alatt. Szótári szó alatt egy olyan valóban létező, vagy mesterségesen konstruált szótövet fogunk érteni, amelyhez toldalékok járulhatnak. Az előző fejezet, amely a morfémák szétválasztásával foglalkozott, részletesen ismertette, hogy milyen toldalékokat akarunk leválasztani, valamint ismertette azt is, hogy milyen törzuszulásokot kezelünk. A szótárba tehát bele kerülnek mindazok a képzett alakok, ahol a képzőt nem vágtuk le, az összetett szavak, és azok a torzult szótövek, amelyeket az előző fázis algoritmikusan nem kezelt. (Pl. a *tesz* tö mellé belekerül a *ten* tö is a *tenni*, *tenném* stb. alakok miatt, és a *tet* tö is a *tettem*, *tetted* stb. alakok miatt. Az igeiktős igéknél az igeiktős tö mint főalak, és az igeiktő és igeiktő nélküli tö mint segédalakok, amiből később össze kell állítani a teljes alakot.)

A szótár minden egyes szóhoz egy sereg információt rendel hozzá. A hozzárendel információ három csoportba oszthatjuk: morfológiai, szintaktikai és szemantikai információkra.

### Morfológiai jellemzők

A morfológiai információ arra a célra szolgál, hogy segítségével megállapítsuk, hogy hol lehet a szóról a végződést leválasztani. Erre nézve több különböző módszer is kialakult. A legtökéletesebb módszer persze az, lenne ha minden szóhoz felsorolhatnánk, hogy milyen végzésekkel kaphat és azokat milyen nyelvtani értelemben. Ez azonban mérhetetlen nagy tömegű információ volna. Ezért az egyes szavak nyilván ragozási osztályokra fognak utalni. Sajnos, elég finom felosztás mellett a ragozási osztályok száma is igen nagy, ezért ezt különböző kivétel-leírási módszerekkel kombinálják. Ezekben azt írják le, hogy a szó bizonyos eseteiben az egyik, más esetekben viszont a másik táblázat érvényes, míg további esetekben explicite megadják a kivételes alakot. Ilyen módszert alkalmazott Ju. Apreszján az ETAP fordító-programoknál.

Egy másik lehetséges megoldás, hogy a szótő tulajdonságait egy tulajdonság vektorral írjuk le (pl. szófaj, magas illetve mély hangrend, ajakkerekítésesség stb.) és egy hasonló vektorral jellemezzük a végződést. A két vektor kölcsönhatásából azután megállapítható, alkalmazható-e a szóra a végződés és milyen morfológiai tulajdonságokkal rendelkezik a keletkezett szó. Ilyen módszert alkalmazott a Kiss Zoltán a KFKI-ban készített gépelést ellenőrző programban.

A mi módszerünk ezeknél sokkal durvább. A szavakat egyszerűen csoportokba osztjuk aszerint, hogy milyen végződéseket kaphatnak. Igeragot igék kaphatnak, birtokos és esetragot névszók, fokozni melléknévszerű szavakat lehet stb. Minket is meglepett, hogy ezzel az egyszerű módszerrel statisztikusan igen jó analízist kaptunk. Természetesen a szó végéről a potenciális végződéseket mindig levágtuk, de azokat a szétvágásokat elvetettük, ahol a végződés és a fő osztályba sorolása között ellentmondás volt. Kevesebb mint fél százalék volt az olyan eset ahol nem odaillő végződést vágtunk le és az algoritmus ezt nem mutatta ki. Ilyenek voltak: a *szinte* szóban a szint szót is megtalálta -e bitokos raggal ( *szintje* helyett). A *másik* szóban a *más* szót vélte megtalálni -ik birtokraggal (a *kezük* szó analógiájára). Valószínű, hogy a hibák egy részét a raghoz kapcsolt szűrő, transzformáló algoritmusok pontosításával meg lehet szüntetni.

Ez a szám igen kicsi, ahhoz képest, hogy egy szövegben kb. 8% morfológiailag kétértelmű szó van. Ennek nagy része természetesen abból fakad, hogy maga a szótő kétértelmű, például: az névelő vagy mutató névmás, *mi* személyes vagy kérdő névmás, *egy* határozatlan névelő vagy számnév. Ez teszi ki körülbelül a kétértelműségek felét. Továbbá az igéknél gyakran nem állapítható meg, hogy múlt idő egyesszám 3. személyről, vagy multidejű melléknévi igenévről van-e szó. Az esetrag előtti é-ről nem állapítható meg, hogy birtokjel vagy birtokos jel. Pl. a *kését* szónál, nem tudjuk, hogy valakinek a kését vagy a késnek a valamijét emlegetjük-e. Mindezek mellett természetesen vannak olyan kétérterműségek is mint pl. a *minden* az összes és mindegyiken értelemben, vagy *érték* főnévként és ragozott igeiként stb. Mindezeket a kétértelműségeket később, a szintaktikus és szemantikus elemzés során kell feloldani, és minden arra mutat, hogy ekkor a durva morfológiai elemzésből származó félreértelmezések is megszűnnek.

### Szintaktikai jellemzők

A szintaktikai jellemzők lényegében azt írják le, hogy a szó a mondatban milyen szerepet tölthet be, hol állhat a mondatban és milyen alakokban. A betölthető szerepeket természetesen alapvetően meghatározza, hogy a szó milyen szófajhoz tartozik és milyen végződéseket vehet fel. Úgy tűnik azonban, hogy a hagyományos nyelvtanokban bevezetett szófaji osztályozás nekünk nem elég finom, mert a mondatban betöltött szerep és a szó toldalékolhatósági típusa nem feltétlenül esik egybe. Ezért, a hagyományos szófaji osztályokat többé-kevésbé megtartottuk, de tovább finomítottuk azokat. Pl. a *muszáj* szót a szótárak tradicionálisan igeiként vagy segédigeként minősítik, és ez utóbbi teljesen megfelel a mondatbeli szerepének. Ezzel szemben toldalékolás szempontjából semmilyen igei tulajdonságot nem mutat. Miután számos ilyen szót találtunk, ezekre a "nem-igei segédige" elnevezést vezettük be.

Próbaképpen, mi körülbelül 120 Kbyte (kb. 50 nyomtatott oldal) szövegből indultunk ki, amelyet az elmúlt 1 év folyamán írtunk, és a természetes nyelvű ember-számítógép kapcsolattal foglalkozott. Ezen próbálgattuk morfológiai elemzőinket és ebből készítettünk egy szótárt. A szótár körülbelül 2500 szótövet tartalmaz (a fenti értelemben). A következő fejezet részletesen ismerteti az általunk bevezetett szófajták meghatározását, a függelékben pedig megadjuk a fenti 2500 szó ezen elvek szerinti besorolását.

Ez a szófaji besorolás tehát kettős arcú; egyrészt megadja, hogy a szó milyen jellegű toldalékokat vehet fel, másrészt megadja, hogy milyen szerepet tölthet be a szintaxisban. Bennünk is felmerült a kérdés meg lehet-e adni ezt a két vonást egymástól függetlenül. Egyelőre úgy látjuk, hogy a két besorolás nem független, és egy ilyen jellegű szófaji felosztás a szintaktikai analízis számára megfelelő alapot biztosít.

Tudjuk, hogy ez a felosztás még nem végleges, csak nagyobb tömegű szöveg szintaktikai elemzése igazolhatja az egyes bevezetett szóosztályok létjogosultságát, illetve szükségessé teheti azok megváltoztatását.

### **Szemantikai jellemzők**

A szó harmadik jellemzője a szemantikai tartalom. Mi itt, ebben a munkában szemantikának azt nevezzük, amire az illető szót használni lehet az adott feladat megoldásában. Lesznek olyan szók, elsősorban igék, amelyek valamilyen tevékenységet indítanak el a számítógépben. A főnevek táblázatokat (adatállományokat) jelölnek, a főnevek és a mellénevek a táblázatok oszlopait illetve sorait jelölik ki stb. Feltételezzük tehát, hogy a szemantikus információ meglehetősen témakörhöz kötött. Nem áll szándékunkban egy univerzális, minden témakörre alkalmazható általános magyar nyelvi szemantikát kidolgozni, és azután valamilyen áttételen keresztül az adott témakörre alkalmazni. Ilyenformán univerzális, témakörtől független szemantikus leírása valószínűleg csak a kötőszavaknak, módosítóknak, segédigéknek stb. lesz.

A fentieknek az a következménye, hogy a szemantikus információt a különböző feladatoknál ki kell cserélni még azoknak a szavaknak a jó részénél is, amelyek közösek két különböző feladat esetén. Ez indokolhatja azt, hogy a szemantikus információkat tartsuk külön szótárban. Ennek további előnye lenne, hogy a munka két különböző fázisában két, egyenként kisebb szótárral dolgozhanánk. Hátránya viszont, hogy párhuzamosan két szótárt kell karbantartani, és könnyen inkonzisztens állapot állhat elő.

## A magyar szavak morfológiai-szintaktikai felosztása

### 0----- Ragozhatatlanok

#### 01--- Határozószók

##### 011-- határozói igenevek

Ide csak a régies *-ván, -vén* képzős alakok kerültek. A *-va, -ve* képzőt a morfológiai elemző levágja.

##### 012-- határozói névmások

Ide kerültek azok a ragozott névmások is, ahol a tö erősen torzult és ezért elemzésük problematikus.

#### 02--- Igekötők

Igekötő minden olyan szócska amelyet, ha az ige előtt áll, egybeírunk vele, ha utána, külön.

#### 03--- Névelők

##### 038-- határozott névelő

##### 039-- határozatlan névelő

#### 04--- Kötőszók

További osztályozása szükséges aszerint, hogy szavakat vagy mondatokat köt össze; ha mondatokat alá vagy mellérendelő mondatokat. Megjelölendő továbbá a kételemű kötőszavak.

#### 05--- Módosító szavak

##### 051-- Igei módosító szavak

Az ige értelmét módosítják: *volna* feltételes mód (hátsó), *hadd* és *hagy* óhajtás, *ne* és *se* tagadás (elől).

##### 052-- Melléknévi (+igei) módosító szavak

A melléknév, melléknév jellegű számnév, melléknévből képzett határozószó és ige jelentését erősíti, gyengíti, tagadja stb. A szó előtt áll.

##### 053-- Számnévi módosító szavak

A konkrét számok értékét bizonytalanítja; a szám előtt áll.

##### 054-- Névutószerű szavak

A névutók kaphatnak formális birtokosragot, ezek nem.

#### 060-- Igeszerű szavak

##### 061-- Nemigei segédigék

Soha nem állnak jelzőként, mindig főnévi igenévvel. A mondat állítmánya a segédige+főnévi igenév. Az alany *-nak* ragot kap. A főnévi igenév ragozódhat, különösen ha nincs ágens kitéve. Múlt ideje a *volt* szóval, jövő ideje a *lesz* szóval, feltételes módja a *lenne* vagy *volna* szóval képződik. Pl. *Pistának tanácsos odamenni, nem volna szabad odamenned*, stb.

##### 062-- Ragozhatatlan főnévi névmások

A személyes névmások alany és tárgyesetei. A többi esetet lásd a névutók ragozásánál.

#### 07--- Egyéb szavak



- 1---- Ige  
Ami igeragokat kaphat.
- 11--- Segédigék  
A segédige ragozódik, az ige (ragozatlan) főnévi igenévvel áll.
- 12--- Féligői segédigék  
Az állítmány ugyanúgy képződik, mint a nem-igei segédigék esetén, de a múlt idő, jövő idő, feltételes és felszólító módja a segédige egyes szám 3.személyű megfelelő alakjával képződik. Pl. *nekem* el kellene mennem .
- 19--- Főnévi igenév (\* van ragozása, amely az igeragozás része)

2----- Névszó

- 201-- nevek  
A nevek személynevek, cégek, hónapok stb nevei. A névszói szerkezetben (amikor nem önmaga a fő főnév) a főnév, illetve az azt megelőző 211 számmal jelölt főnévi jelző előtt, a tulajdonságjelzők után szoktak állni. Pl. *Jan Kowalsky lenyel vendégmunkás, Kiss Péter lakatos, Pista bácsi, Vörös Október gépgyár, március hónap* stb.
- 202-- mértékegységek
- 21--- főnév
- 211-- egyszerre főnévi és melléknévi szó  
Idetartoznak az anyagnevek, foglalkozások, nemzetiségek, vallások stb. Egyformán gyakran állnak főnévként és közvetlenül a főnév előtt álló jelzőként. Nem fokozhatók. Pl. *arany virágcserep, lengyel vendégmunkás, csőszerelő szakmunkás, katolikus pap, elnök elvtárs* stb.
- 212-- teljesen főnév jellegű névmás  
Teljesen úgy ragozódik, mint a főnevek; birtokos és esetragot is kaphatnak.
- 213-- névutóhoz hasonlóan ragozódó névmások  
A névmás különböző személyű alakjait egy konstruált többől személyragokkal képezhetjük, utána birtokosjel és esetrag is állhat. A személyragok formailag a birtokos ragozás ragjaival esnek egybe. Pl. *magam, magad, maga, magáé, magunktól* stb.
- 214-- birtokosan nem ragozódó névmások  
Birtokjelet nem kaphatnak, de birtokos jelet, és esetragot kaphatnak. Pl. *egymás, egymását, egymást* stb.
- 22--- Melléknév  
Állhat tulajdonságjelzőként névszói szerkezetben, nem-igei segédigeként, nem igei állítmánytként. Megkülönböztetendők azok, amelyek vonzattal rendelkeznek. Fokozhatók.
- 221-- melléknévi igenév
- 222-- melléknév jellegű névmás
- 223-- melléknévi névutó

23--- Számnév

230-- konkrét szám

Egytől tízig betűvel és számmal, továbbá száz, ezer, millió betűvel is, a többi csak számmal, kötőjellel toldalékolva.

231-- melléknév jellegű számnév

Melléknévként viselkedik, fokozható, melléknévi módosítót kaphat, de a névszói szerkezetben számnév helyén áll a tulajdonság jelzők előtt. A -szor raggal számhatározó képezhető belőle.

232-- névmás jellegű számnév

Nem mutat melléknévi tulajdonságokat, a számjelző helyén állhat. Számhatározó képezhető belőle. Számnévi módosítót kaphat. Nagyrészüik számnévi névmás.

233-- kvantor jellegű

A névszói szerkezetben a számjelző előtt az un. kvantor pozícióban áll. Jöhet utána számjelző is. Számhatározó nem képezhető belőle.

24--- Névutó

A személyes névmások ragozott illetve névutóval ellátott alakját, a névutó+személyrag alakban képezzük. A bizonyos ragoknak egy mesterséges fő felel meg, pl. a -val -vel ragnak a vel- : velem, veled, vele; az -on -en -ön ragnak a rajt- : rajtam, rajtad, rajta stb. Megkülönböztetendőek azok a névutók, amelyek ragot vonzanak.

Egyéb jelölések:

----1 nem szótári alak, csak ragozott formában szerepelhet

----2 ragozott alak (a fő torzulása miatt került a szótárba)

----4 nem kell tovább bontani, csak ez a forma szerepel a szótárban

---1- kötelező vonzata van

---2- opcionális vonzata van

## A szintaxis alapjai

A szintaxis a mondatok felépítésével foglalkozik. A mondatban a szavak bizonyos egységekké kapcsolódnak össze, melyek azután ismét nagyobb egységeket alkotnak, és ezekből még tovább; amíg a teljes mondat fel nem épül. A magyar nyelv, mint az közzismert, ún. szabad szórendű nyelv. Ez természetesen nem azt jelenti, hogy a magyar mondatban a szavak tetszőleges sorrendben állhatnak. A magyar mondatban a szintagmákat (mondattani egységeket) két kategóriába lehet sorolni. A felsőbb szintű mondattani egységekre az a jellemző, hogy mondatbeli sorrendjük szabad, és az egyes egységek grammatikai szerepét az egység utolsó tagjaként álló morféma (rendszerint egy rag vagy egy névutó) szabja meg. Az alsóbb szinten viszont a mondatrész alkotórészeinek egymáshoz kapcsolódását pozícionális, nevezetesen sorrendi szabályok írják elő.

A szabad szórend nem jelenti azt sem, hogy a felsőbb szinten a szintagmák sorrendje tetszőleges. Ezen a szinten a sorrendet nem a mondattani szabályok írják elő, hanem a beszélő szándéka. A kérdő, tagadó, felszólító mondatokban a sorrend valamivel kötöttebb, a kijelentő mondatokban pedig nagyrészt attól függ, hogy mit kívánunk kihangsúlyozni.

Ezt a jelenséget É.Kiss Katalin igen részletesen leírta, mi ebben a fejezetben ezzel a sorrendi kérdéssel nem akarunk foglalkozni, de később még visszatérünk rá.

Mint más nyelvekben is, a magyarban is a mondat fő alkotó eleme az ige, pontosabban az igei kifejezés; ez akkor is igaz, ha a mondatban az ige nincs is explicite kitéve. Az ige jelentése szabja meg, hogy milyen egyéb mondatrészek szerepelhetnek még a mondatban. Ezeket az igehez kapcsolódó mondatrészeket az ige argumentumainak fogjuk nevezni. Lehet az ige logikai alanya aki (vagy ami) a cselekvést végrehajtja, vagy akivel valami történt (pl. *alszik+valaki*, *esik+valaki*); lehet ezenkívül tárgya (pl. *olvas+valaki+valamit*) vagy részeshatározója (kedvezményezettje) (pl. *ad+valaki+valamit+valakinek*). Ezen argumentumokat többnyire névszói kifejezések adják meg, és az egyes argumentumokat a kifejezés végén álló rag vagy névutó jelöli ki.

Az az absztrakt szemléletmód, hogy a mondat egyes alkotórészeit funkciójuk, és nem nyelvtani esetük szerint különböztetjük meg, fontosnak bizonyult a gépi fordítás esetén. Ugyanis különböző nyelvekben az egyes funkciók különböző névszói szerkezetekben jelennek meg. De azt is látni fogjuk, hogy magán a magyar nyelven belül is részben az igei kifejezés struktúrájától függ, hogy egy argumentum milyen formában kerül a mondatba. A legegyszerűbb esetben az logikai alany a mondat alanya, de a magyar nyelvben sok olyan konstrukció is van a mikor az logikai alany -NAK ragot kap.



A szintaxison belül először az igei szerkezetek felépítésével foglalkozunk. A következő részben az argumentumként szolgáló névszói szerkezetek felépítését ismertetjük. Ezek után a két legfontosabb argumentumnak az alanynak és a tárgynak a felismerésével foglalkozunk az ige-alany és az ige-tárgy egyeztetések kapcsán. Végül a "Vonzatok" című részben arról lesz szó, hogyan ismerjük fel az ige további argumentumait az esetragok és névutók segítségével.

### Igei szerkezetek

Az egyszerű (nem összetett) magyar mondatok az ige szempontjából a következő csoportokba sorolhatók:

1. Ragozott ige, a logikai alany alanyesetben van, az igenek néhány kötelező és néhány opcionális argumentuma van. Pl. *Pista olvas.*

2. Igei ragozású segédige, mellette az ige főnévi igeneves formában, a kettőnek azonos logikai alanya van alanyesetben. Pl. *Pista olvasni fog.*

3. Nem-igei segédige, mellette az ige infinitívusban vagy infinitívuszi ragozásban, a logikai alany -NAK ragot kap. Pl. *Pistának olvasnia kell.*

4. A VAN ige "létezik" értelemben. Az egy argumentuma van alanyesetben. Pl. *Hideg van.*

5. A VAN ige "birtokol" értelemben. A logikai alany -NAK ragot kap, amit birtokol, alanyesetben áll birtok jellel. Pl. *Pistának van mai újsága.*

6. Szenvedő szerkezet. A tárgy van alanyesetben, az ige VAN+határozói igenév alakban áll, a logikai alany, ha egyáltalán meg van adva, által névutót kap. Pl. *A feladat meg lesz oldva.*

7. Valami valamilyen, vagy valami valahány. A jelen idejű VAN nincs kitéve. Mindkét argumentum alanyesetű. Az első egy névelős névszói kifejezés, (személynevek elé nem kerül névelő), a másik egy melléknévi vagy számnévi kifejezés (nem lehet névelője). A sorrend tetszőleges. Pl. *A rózsza piros volt. Sok volt az eső. Az út 3 km volt.*

8. Valami valami. ("is-a" kapcsolat) A jelen idejű VAN nincs kitéve. Mindkét argumentum alanyesetű főnévi kifejezés. A sorrend lényeges és a mondat típusától (kijelentő, kérdő, tagadó) függ. Pl. *Pista katona volt. A katona Pista volt.*

A mondat központjában álló igei kifejezés meghatározza (az igeen kívül) az ige módját, idejét, logikai alanyát és egy sor más vonást. Az igei kifejezés a ragozott igeen kívül, segédigéből, igei módosítóból, szabad határozókból áll, és

explicite megadott logikai alanya is lehet. Ezek együtt adják meg a fenti jellemzőket.

Az igei kifejezés bizonyos mértékig mindig meghatározza az ige alanyát is. 1. és 2. személyben nem tekinthető hiányosnak a magyar mondat, ha az alany csak az igeraggal van megadva. Ha a névmás is ki van téve, a forma redundáns, és az ismételt utalás csak a hangsúlyozás miatt van. A 3. személy esetén a névszói kifejezéssel megadott alany hiányát utalásnak kell tekinteni (általában hátra, néha azonban előre). Az igerag által indikált grammatikai alanynak és a mondatban explicite megadott alanynak bizonyos vonásokban meg kell egyezniük. Ezt az "Egyeztetések" című fejezet definiálja pontosabban.

A grammatikai alanyt a személyrag adja meg. Ezt az 1., 4., 7. és 8. esetben az ige kapja, a 2. esetben a segédige. A 3. esetben az infinitívusz vagy a *nek* ragból képzett névmás vagy mindkettő; az 5. esetben ugyanezen a névmás illetve a birokjel adja meg.

A 6. esetben, ha nincs kitéve, hogy ki által, és a 2. esetben, ha se az infinitívusz nem ragozódik, se névmás nincsen, általános alanyról beszélhetünk. Pl. *A probléma meg lesz oldva. Tilos bemenni.*

Az ige módján általánosabb értelemben azt értjük, hogy a cselekvés feltételes, lehetséges, szükségszerű, rendszeres, stb. Ez a különböző igei kifejezésekben, különböző formákban jelenik meg. Az ige általánosabb értelemben vett módját megadhatja részben az igerag, részben igei képző (pl. a *-hat, -het*), részben segédige, részben az igei módosítók (pl. *volna*). Pl. *olvasna, tudna olvasni, olvashat, olvashatna, kellene olvasnia, muszáj lenne olvasnia, hideg volna, kell lennie mai újságjának, katona volna* stb.

Hasonló a helyzet az igeidővel is. Az időt megadhatja rag, segédige, segédige ragja, igei módosító. Pl. *olvasta, olvasni fogja, kellett olvasnia, muszáj lesz elmennie* stb. De megadhatja az ige idejét, úgynevezett szabad határozó is. Pl. *Holnap elolvassa.*

Összefoglalva: a mondat központjában az ige áll. Az igenek attribútumai vannak, amelyek megadják az ige lefolyásával kapcsolatos tudnivalókat időt, helyet, alanyt, befejezettséget, feltételességet, lehetségességet, ismétlődést, stb. Ezeket az attribútumokat, megadhatják igeragok, segédigék, igei módosítók, képzők és a mondat szabad határozói. Azt, hogy a rendszer ezen tulajdonságok közül melyeket és milyen osztályozásban tartja számon, a rendszer céljaitól függ.

Célszerűnek látszik a következő attribútumokat felvenni: alany, idő, hely, mód. Az attribútumok értéküket egy adott értékkészletből veszik. Az attribútumok értéke lehet definiálatlan is; ez esetben a balodalon az attribútum default értéket kap.

# Formális definíció

A fentiek szerint a mondat felépítése a következő attribútum nyelvtannal megadva:

(Jelölések:

```

::=      definiáló egyenlőség a jobb oldal komponenseinek
          sorrendje között
+        ezen két komponens között a sorrend kötött
mondat   grammatikai fogalom
(...)    a fogalom által definiált attribútumok
VAN, NAK  explicite megadott morféma
          (a mondatban helyette a létige valamelyik alakja vagy
          -nak illetve -nek rag áll.)
[...]    opcionális )
  
```

(Megjegyzés: a formális definícióban helytakatékosság céljából nincs leírva, a

határozói igenév ::= ige+VA

és más hasonló magától értetődő szabályok.)

mondat ::= igei kifejezés(alany,hely,idő,mód)

```

igei kifejezés ::= [főnévi kifejezés(alany)]
                  ige+IGEI KÉPZŐ(mód)+IGERAG(alany,idő,mód)
                  [hol(hely)] [mikor(idő)]
  
```

```

Pl. Pista      tegnap    a kertben  olvas-hat      -ott.
    (alany=    (idő=      (hely=      (mód=      (alany=
    e.sz.3.sz) mult)      adott)      lehet.)    e.sz.3.sz,
                                   idő=
                                   mult,
                                   mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott,hely=adott,idő=mult,
                                                         mód=lehetséges)
  
```

```

Holnap      elolvas-om.
(idő=      (alany=
jövő)      e.sz.1.sz.,
           idő=0,
           mód=0)
  
```

igei kifejezés(alany=e.sz.1.sz.,hely=0,idő=jövő,mód=kijelentő)

Nem mondat ezzel szemben a következő:

```

Pl. *Pista iskolába olvasott.
    *Pista holnap olvasott.
  
```

```
igei kifejezés2 ::= [főnévi kifejezés(alany)]
                    segédige+IGERAG(alany,idő,mód)
                    főnévi igenév
                    [hol(hely)] [mikor(idő)]
```

```
Pl. Holnap      fog olvasni.
    (idő=       (alany=
    jövő)      e.sz.3.sz.,
               idő=
               jövő,
               mód=0)
```

```
igei kifejezés(alany=e.sz.3.sz.,hely=0,idő=jövő,mód=kijelentő)
```

```
igei kifejezés3 ::= [főnévi kifejezés(alany)+NAK]
                    segédige+IGERAG(idő,mód)
                    főnévi igenév+[IGERAG(alany)]
                    [hol(hely)] [mikor(idő)]
```

```
Pl. Nekem      kell-ett      elmen-nem.
    (alany=     (idő=         (alany=
    e.sz.1.sz.) mult,         e.sz.1.sz.)
               mód=0)
```

```
igei kifejezés(alany=e.sz.1.sz.,hely=0,idő=mult,mód=szükséges)
```

```
El kell-ene      men-nünk.
    (idő=0,      (alany=
    mód=         t.sz.1.sz)
    felt.)
```

```
igei kifejezés(alany=t.sz.1.sz.,hely=0,idő=jelen,
               mód=feltételes/szükséges)
```

```
igei kifejezés3 ::= [főnévi kifejezés(alany)+NAK]
                    segédige VAN(idő,mód)
                    főnévi igenév+[IGERAG(alany)]
                    [hol(hely)] [mikor(idő)]
```

```
Pl. Tilos      volt      bemenni.
    (mód=       (idő=
    tagadó)     mult)
```

```
igei kifejezés(alany=általános,hely=0,idő=mult,
               mód=kijelentő\tagadó)
```

```
Pl. Muszáj      bemen-nem.
    (mód=         (alany=
    szükséges)    e.sz.1.sz)
```

```
igei kifejezés(alany=e.sz.1.sz.,hely=0,idő=jelen,mód=szükséges)
```

```
igei kifejezés4 ::= [főnévi kifejezés(alany)]
                    VAN(alany=3.sz,idő,mód)
                    [hol(hely)] [mikor(idő)]
```

Pl. *Irak-ban háború van.*  
 (hely= (alany= (alany=  
 adott) e.sz.3.sz) e.sz.3.sz,  
 idő=0,  
 mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott,hely=adott,idő=jelen,  
 mód=kijelentő)

*Holnap le-het -nek zivatarok.*  
 (idő= (mód= (alany= (alany=  
 jövő) lehet.) t.sz.3.sz, t.sz.3.sz)  
 idő=0,  
 mód=0)

igei kifejezés(alany=e.sz.3.sz.\adott,hely=0,idő=jövő,  
 mód=lehetséges)

igei kifejezés5 ::= [főnévi kifejezés(alany)+NAK]  
 VAN(idő,mód)  
 névelőtlen főnévi kifejezés+BIRTOKRAG(alany)  
 [hol(hely)] [mikor(idő)]

Pl. *Új bicikli-d van?*  
 (alany= (idő=  
 e.sz.2.sz.) jelen,  
 mód=  
 kijelentő)

igei kifejezés(alany=e.sz.2.sz.\adott,hely=0,idő=jelen,  
 mód=kijelentő)

igei kifejezés6 ::= [főnévi kifejezés(object)]  
 VAN(idő,mód)  
 határozói igenév  
 [hol(hely)] [mikor(idő)]

Pl. *Itt van a kutya elásva.*  
 (hely= (idő= (object=  
 adott jelen, adott\határozott)  
 mód=0)

igei kifejezés(alany=általános,hely=adott,idő=jelen,  
 mód=kijelentő)

(Megjegyzés: Az igék között vannak tárgyas és tárgyatlan igék. Esetenként tárgyatlan igékből is képezhető ilyen "szenvendő" szerkezet. Az ilyen igéknél a szerkezet nyelvtani alanya azonos a logikai alannyal, hasonlóan ahhoz, hogy a múlt idejű és jövő idejű melléknévi igenév után álló jelzett szó sem a jelző tárgya hanem az alanya. Pl. *A csésze össze lett törve. A ruha el van szakadva. Meg van halva. Szépen meg van fésülködve.* stb.)

Vannak azonban olyan tárgyatlan igék is, ahol ilyen szerkezet nem képezhető sőt, esetenként a múlt idejű és jövő idejű melléknévi igenév sem. Pl. *törekszik, rájön, függ, árt.*)



```
igei kifejezés7 ::= [főnévi kifejezés(alany)]
                    melléknévi kifejezés+[VAN(alany,idő,mód)]
                    [hol(hely)] [mikor(idő)]
```

P1. *Vasárnap szép lesz az idő.*

```
(idő=      (alany=      (alany=
jövőimult) e.sz.3.sz,    e.sz.3.sz)
            idő=
            jövő,
            mód=0)
```

```
igei kifejezés(alany=e.sz.3.sz.\adott,hely=0,idő=jövő,
                                                         mód=kijelentő)
```

```
igei kifejezés7 ::= [főnévi kifejezés(alany)]
                    számnévi kifejezés+[VAN(alany,idő,mód)]
                    [hol(hely)] [mikor(idő)]
```

P1. *125 kg vagyok.*

```
(alany=
e.sz.1.sz.,
idő=
jelen,
mód=0)
```

```
igei kifejezés(alany=e.sz.1.sz.,hely=0,idő=jelen,mód=kijelentő)
```

```
igei kifejezés8 ::= [főnévi kifejezés(alany)]
                    főnévi kifejezés+[VAN(alany,idő,mód)]
                    [hol(hely)] [mikor(idő)]
```

P1. *Az igazgató vagyok.*

```
(alany=
e.sz.1.sz.,
idő=0,
mód=0)
```

```
igei kifejezés(alany=e.sz.1.sz.,hely=0,idő=jelen,mód=kijelentő)
```

(Megjegyzés: ha ki van téve az ige, akkor az előtte levő szó tartozik az állítmányba. Ha nincs kitéve, és az egyik határozott, a másik nem, akkor határozott az alany, máskülönben az a feltételezés, hogy az első az alany, a második az állítmány.)

P1. *Pista katona.* (Pista határozott, mert személynév.)

*Ócskavas ez az autó.*

*A lengyel a katona.*

## A névszói kifejezés

Az igei kifejezésben álló igehez mondatrészek kapcsolódnak. Ezek a mondatrészek leírhatják az ige eszközt, tárgyát, helyét, irányát, stb. Alapjában véve úgy képzelhetjük, hogy az ige egy függvény, amelynek meghatározott számú argumentuma van. Természetesen az argumentumok száma és típusa igénként más és más. Pl. *esik* (nincs argumentum), *alszik+ki*, *lát+ki+mit*, *összeköt+ki+mit+mivel*, *odaér+ki+hova*, *megkülönböztet+ki+miket*, *megjelöl+ki+mit+mivel* stb. Az igeeknek ezeket az argumentumait az jelöli ki, hogy a hozzátartozó névszói szócsoporthoz (szintagma) milyen végződéseket vagy névutókat kaphat, illetve hogy milyen határozószók és határozói névmások tartoznak bele). Egyes argumentumok főnévi igenevet vagy *hogy* kötőszós mellékmondatot igényelnek vagy ezek közül valamelyiket.

A névszói argumentum fő eleme általában egy főnév, amelyhez különböző jelzők és esetleg egyéb bővítmények kapcsolódhatnak. Ha a főnév hiányzik, ami a magyar mondatokban igen gyakori, a rag (illetve névutó) az előtte álló szóhoz kapcsolódik.

A jelzők kétféleképpen kapcsolódhatnak a főnévhez. Van olyan eset, amikor a kapcsolódásra csak az utal, hogy a jelző a főnév előtt áll. *Som* milyen nemben, számban, esetben történő egyeztetés nincs, nem úgy, mint más nyelvekben. Ilyenkor a jelzők sorrendje kötött: birtokos jelző, (névelő,) számjelző, tulajdonságjelző, főnévi jelzők, főnév; természetesen bármelyikük el is maradhat. Pl. *a megye három élenjáró Arany Kalász szövetkezete*.

Ha a főnévnek van birtokos jelzője, akkor a fenti esetben a birtokos nem kap semmilyen végződést, (és a jelzett szó vagy kifejezés nem kaphat határozott névelőt,) a birtokolt főnév pedig birtokosragozásbeli birtokjelet kap.

Van a birtokviszonynak egy másik alakja is, ekkor a birtokos *-nak -nek* ragot kap, a birtok és jelzői elé viszont (többnyire határozott) névelő kerül. Ez esetben viszont a birtokos és a birtok el is szakadhat egymástól a mondatban. Pl. *Pistának látom az árnyékát*.

Nemcsak a birtokos jelző hanem bármelyik jelző elszakadhat a főnév mellől és ún. hátravetett jelző lehet belőle. Ilyenkor viszont ugyanolyan ragot kap, mint a jelzett szó. A (-NAK ragtalan) birtokos jelzőt is hátra lehet vetni és ekkor a főnév többnyire nem kap birtok ragot, viszont a birtokos megkapja a birtoklás *-é* vagy *-éi* jelét. Pl. *Kalapot vettem tegnap a vásárban, pirosat. Elvesztettem a levelet, Jenőét*.

Ha a birtokos (határozott névelős) személyes névmás, az ezzel egyenértékű hátravetett jelző a birtokos névmás lesz. Pl. *az én kalapomat..., a kalapot, az enyémet ...* A birtokos névmás csak olyan helyen állhat, ahol a valódi argumentumként szereplő birtokot nem tettük ki.



Ha az argumentumként álló főnév vagy annak valamilyen jelzője igei jelentésű, akkor ahhoz is kapcsolódhatnak tárgy, határozó és más igei jellegű bővítmények. Pl. *iskolába járás során*. Néhány más szónak, pl. melléknévnek is lehet argumentuma. Pl. *valamilyen+színű*.

A jelzők maguk is lehet jelzős kifejezések. Pl. *a téesz bikájának a szarva, múlt idejű ige*.

Argumentumként több kifejezés is állhat *és, vagy, vagy... vagy..., nem csak ... hanem .... (is)* és más hasonló szavakkal összekapcsolva.

Ha névmás áll az argumentum pozíciójában, akkor a jelző csak hátravetett lehet. Pl *Ő, a katona elaludt. Az enyémet, a kéket nem adom*.

#### Formális leírás

argumentum1 ::= névszói kifejezés+[ESETRAG]+[NÉVUTÓ]

(Megjegyzés: vannak névutók, amelyek esetraggal járnak. A birtokrag után viszont nem feletétlenül tesszük ki a tárgyragot.

Megjegyezzük továbbá, hogy az ige argumentumainak vonzatát nem befolyásolja az, hogy milyen igei kifejezéshez kapcsolódnak, van-e segédige vagy se, kivéve a szenvedő szerkezetet, amikor a szerkezet alanya valójában tárgy. Segédigének pont azokat az igéket nevezzük, amelyek szükségszerűen infinitívussal járnak és az infinitívusban álló ige vonzatait nem változtatják meg.)

argumentum2 ::= határozószó

argumentum3 ::= határozói névmás

argumentum4 ::= névmás+[TÁRGYRAG]

(Megjegyzés: az ige egyetlen logikai argumentumához, mint azt fentebb vázoltuk több argumentum jellegű szócsoporthoz kapcsolódhat; ezek összetartozásának megállapítása nem könnyű és nem is mindig megoldható feladat, de a gyakorlatban a kétértelmű mondatok nem túl gyakoriak. A következő feltételezésekkel élünk:

1. Ha sem az igének, sem a segédigének nincs *NAK* ragos vonzata és van egy *NAK* ragos kifejezésünk, és egy másik, birtokos nélküli, bírókjellel ellátott névszói kifejezés, akkor felételezhetjük, hogy az előbbi az utóbbi birtokosa.

2. Ha van egy névszói kifejezés, és a mondatban hátrébb egy ugyanolyan esetraggal ellátott, birtokosjelet viselő, tehát hiányzó birtokú névszói kifejezés, felételezhetjük, hogy az előbbi az utóbbi birtoka.

3. Ha van két azonos raggal vagy névutóval ellátott argumentum és az utóbbiból hiányzik a főnév, akkor feltétlezhetjük, hogy az utóbbi az előbbinek hátravetett jelzője.

Az előbb felsoroltak mindegyikére lehetne valószínűleg ellenpéldát konstruálni, így ezeket is érdemes szemantikai jegyek alapján ellenőrizni. Ellenkező esetben több lehetséges szintaktikai struktúrát építhetünk fel, amelyek közül válogatnunk kell szemantikai jegyek alapján. Bizonyos esetekben azonban az argumentum hovatartozása nem lesz megállapítható. Különösen a NAK rag okoz problémát amelyet számtalan különböző célra használunk. Pl. *Péternek kell elnevezni.* (Az *elnevező* vagy az *elnevezett Péter?*)

névszói kifejezés1 ::= birtokos szerkezet

névszói kifejezés2 ::= névelős kifejezés

névelős kifejezés1 -> AZ+névelőtlen kifejezés

névelős kifejezés2 -> határozatlan kifejezés

határozatlan kifejezés -> [EGY] névelőtlen kifejezés

birtokos szerkezet1 ::= nak ragos birtokviszony

birtokos szerkezet2 ::= egybefüggő birtokviszony

birtokos szerkezet3 ::= hiányos birtokviszony

nak ragos birtokviszony ::= [névelős kifejezés(tulaj.)+NAK]  
névelős kifejezés+BIRTOKJEL(tulaj.)

egybefüggő birtokviszony ::=névelős kifejezés(tulaj.)  
határozatlan kifejezés+BIRTOKJEL(tulaj.)

hiányos birtokviszony ::= névelős kifejezés+BIRTOKOSJEL

névelőtlen kifejezés1 ::= [számjelző]+[tulajdonság jelző]+  
+[főnévi jelző]+[főnév]

(Megjegyzés: természetesen a 4 rész mindegyike nem lehet üres.)

névelőtlen kifejezés2 ::= főnévi névmás

(Megjegyzés: általános szabály, hogy egy adott szinten lévő névmást nem előzhet meg semmi. Tehát, ha a főnév helyén áll névmás nem lehet jelzője, a tulajdonságjelzők közül mindig az első, ha számnév nem lehet kvantora. Pl. *\*piros ez, olyan nagyfogú fésű, \*nagyfogú olyan fésű, annyi méterre, minden 3 métrre, \*minden annyi méterre.* De a különböző szintek nem zavarják egymást, pl. *3 olyan fésű.*)

(Megjegyzés: a meghatározás után álló (xxx--) alakú szám, a szó szóosztályára utal.)

főnévi jelző ::= [tulajdon név (201--)]+  
                  +[egyszerre főnév és melléknév (211--)]

számjelző1 ::= [kvantor (233--)]+  
                  +[számnévi módosító (053--)]+  
                  +[konkrét szám (230--)]+  
                  +[mértékegység]

számjelző2 ::= [számnévi névmás (233--)]+[mértékegység]

számjelző3 ::= [melléknévi módosító (052--)]+  
                  +[melléknév jellegű számnév (232--)]+  
                  +[mértékegység]

tulajdonság jelző1 ::= melléknévi névmás (222--)

tulajdonság jelző2 ::= [tulajdonság jelző]+melléknévi szerkezet

tulajdonság jelző3 ::= [tulajdonság jelző]+  
                          +melléknévi igeneves szerkezet

tulajdonság jelző4 ::= [tulajdonság jelző]+  
                          +melléknévi névutós szerkezet

melléknévi szerkezet ::= [melléknévi módosító]+melléknév

melléknévi igeneves szerkezet ::= [melléknévi módosító]+  
  +[argumentumok és határozók]+  
  +melléknévi igenév

melléknévi névutós szerkezet ::= névszói kifejezés+[ESETRAG]+  
  +MELLÉKNÉVI NÉVUTÓ(223--)

argumentumok és határozók ::= [argumentum] ... [argumentum]  
                                  [hol][mikor][módhatározó]

(Tapasztalatunk szerint a melléknévi igenév mellett  
leggyakrabban módhatározó áll).

### Az alany, a tárgy és az ige

A magyar nyelvben az igének az igerag mindig megadja a grammatikai alanyát, és a tárgyas igék esetében a tárgyat is. Ez a következő lehet:

egyes szám 1. személy= a beszélő;

egyes szám 2. személy= a hallgatója, vagy egy a hallgatók közül;

egyes szám 3. személy= valaki más, vagy valamilyen más tárgy  
vagy fogalom;

többes szám 1. személy= a beszélő és a hallgató vagy hallgatók;

többes szám 2. személy= a hallgatók;

többes szám 3. személy= más valakik vagy más tárgyak  
vagy fogalmak.

A grammatikai alanyon és tárgyon kívül a mondatban lehet explicit névszói kifejezéssel vagy névmással megadott alany és tárgy is. Ha van explicite megadott alany vagy tárgy, annak meg kell egyeznie a grammatikai alannyal. Az alábbi szabályok azt foglalják össze, hogy mikor kell kirakni az explicit alanyt vagy tárgyat. Ha kellene, de még sincs kitéve, azt utalásnak fogjuk tekinteni, amiről meg kell állapítanunk, hogy mire utal.

Az első és a második személy mindig jól meghatározott, anélkül, hogy bármi egyebet tudnunk kellene. Nem ez a helyzet a harmadik személlyel, akinek vagy aminek a pontosabb meghatározásához további információra van szükség.

A magyar nyelv általában kerüli a redundanciát és ezért az első és második személyben a névmási alanyt nem tesszük ki, hacsak nem akarjuk hangsúlyozni.

A harmadik személyben viszont általában kitesszük az alanyt a pontosítás miatt. Ha a harmadik személyű alany nincs kitéve, akkor ez rendszerint hivatkozás valamire vagy valakire, akit korábban a diskurzus fókuszába helyeztünk. Nem kell megadni az alanyt, ha az magától értetődő, pl. *Villámlik.*

Az igéknek van tárgyas és tárgyatlan ragozása. Nagyon kevés olyan ige van amelyik alkalmilag sem áll tárgyas ragozásban. Pl. *lefutja a napi 10 kilométert, az igazak álmát aludta,* stb. Az ige tárgyatlansága, nem a tárgy hiányából látszik sokkal inkább abból, hogy múlt idejű melléknévi igeneve cselekvő.

A magyar igeragozásban a tárgyra a következő lehetőségek vannak: meghatározatlan, 1.személy, 2.személy, 3.személy. Látható, hogy a tárgy egyes vagy többes száma az igeragozást nem befolyásolja.

tárgy	határozatlan	1.személy	2.személy	3.személy
alany				
e.sz.1.sz	A:látok valakit	T:látom magamat	?:látlak téged	T:látom őt
e.sz.2.sz	A:látsz valakit	A:látsz engem	T:látod magadat	T:látod őt
e.sz.3.sz	A:lát valakit	A:lát engem	A:lát téged	T:látja őt/magát
t.sz.1.sz	A:látunk valakit	T:látjuk magunkat	A:látunk téged	T:látjuk őt
t.sz.2.sz	A:láttok valakit	A:láttok engem	T:látjátok magatokat	T:látjátok őt
t.sz.3.sz	A:látnak valakit	A:látnak engem	A:látnak téged	T:látják őt/magukat

Tradicionálisan az első oszlopot nevezik alanyi, az utolsót tárgyas igeragozásnak. Mivel az ige alanyának és tárgyának egybeesése viszonylag ritka, ezt explicite meg kell adni. Ezeket az eseteket elhagyva a következő szabályok alakulnak ki:

1. A tárgyas ragozás mindig harmadik személyű tárgyat vonz. Mivel a harmadik személyű tárgy nincs pontosan meghatározva, ezért megadása szükséges, illetve, ha hiányzik hivatkozásnak tekintjük.

2. Ha az alanyi ragozás egyes szám első személyű alakjánál hiányzik a tárgy és az igének kötelező vonzata a tárgy, az utalás valami határozatlanra.

3. Ha az alanyi ragozás egyes szám második vagy harmadik személyű, illetve többes szám második vagy harmadik személyű alakjánál hiányzik a tárgy, feltételezzük az *engem* vagy *minket* névmást. Pl. *Szeretsz?*

4. Ha az alanyi ragozás többes szám első személyű alakjánál hiányzik a tárgy, feltételezzük a *téged* vagy *titeket* névmást. Pl. *Megkeresünk.*

5. Ha az alany egyes szám első személyű és a tárgy második személyű a *téged* szót nem szokás kirakni, és néha a *titeket* se teszik ki. Viszont a főnévvel megadatott tárgy nem kap ragot. Pl. *Szeretlek, Éva.*

Ha a tárgy explicite is meg van adva, annak egyeznie kell tárgyasság tekintetében az igeraggal. Határozott a tárgy, ha van határozott névelője, ha tulajdonnév, ha birtokos vagy birtokjelet visel, illetve ha maga egy olyan névmás, amely tárgyas ragozást vonz és nem tűr maga előtt határozott névelőt, vagy olyan névszói kifejezés, amelynek élén ilyen névmás áll.

(Megjegyzés: az igereszerűen ragozott segédigék esetén a segédige kerül tárgyas ragozásba. Pl. *Meg akarom a feladatot oldani.*)



Pl. *Nem olvastad a könyvet?*  
*Pistának nem olvastad a könyvét?*  
*Nem olvastad a könyvét?*  
*Pistának még nem olvastad könyvét?* ("Nem tudom van-e neki.")  
*Melyik könyvet olvastad?*  
*Mindkét könyvet olvastad?*  
*Azt a könyvet olvastad?*  
*Azt olvastad?*

Határozatlan a tárgy, ha nem határozott. A valamennyi névmás mindkét csoportba beletartozik, de tárgyas ragozás esetén az "összeset" jelenti, tárgyatlan esetben "néhányat". Pl. *Olvastad valamennyit? Olvastál valamennyit?*

Egyes szám harmadik személy kivételével a birtokjel után nem kötelező kitenni a tárgyragot. Pl. *Vettem a kalapom.*

Az alany illetve a tárgy nemcsak egyetlen valami lehet, hanem felsorolás vagy "és"-kapcsolat is. Ebben az esetben a személy és szám a következőképpen alakul:

én+te=én+ő=	én vagy mi	<i>Én és Pista elmegyek oda.</i> <i>Én és Pista elmegyünk oda.</i>
én+ti=én+ők=	mi	<i>Én és Pistáék elmegyünk oda.</i>
te+ő=	te vagy ti	<i>Te és Pista elmész oda.</i> <i>Te és Pista elmentek oda.</i>
te+ők=	ti	<i>Te és Pistáék elmentek oda.</i>
ő+ő=	ő vagy ők	<i>Ő és Pista elmegy oda.</i> <i>Ő és Pista elmennek oda.</i>
mi+te=mi+ő=mi+ti=mi+ők= mi		<i>Mi és Pista elmegyünk oda.</i>
ti+ő=ti+ők	= ti	<i>Ti és Pista elmentek oda.</i>

A számbeli egyeztetés kapcsán meg kell jegyeznünk, hogy számnevek után mindig egyes számot használunk, de csak a mondaton belül. Ha következő mondat ugyanezekre hivatkozik (ki nem tett explicit alannyal), akkor rendszerint már többes számot használunk. Pl. *A két gyerek elment hazulról. A mozi előtt találkoztak.* Hasonló a helyzet a csoportos jelentésű főnevekkel. Pl. *Osztályunk elment moziba.* A 6. és 7. sorban ültünk. Tehát megkülönböztetendő a mondatrész grammatikai száma (és személye) és a logikai száma (és személye) amivel később hivatkozhatunk rá.

(Megjegyzés: az aki névmásnak nincs személye, minden személyraghoz alkalmazható. Pl. *Én voltam az, aki megtudtam a dolgot.*)

A fentiekből látható, hogy az ige egyeztetése az alannyal és a tárggyal közepes méretű feladatnak látszik. Felmerül a kérdés, hogy szükséges-e ezt a feladatot az elemzés során elvégezni. Mivel nem az a célunk, hogy szöveg nyelvtani helyességét ellenőrizzük, és a magyar anyanyelvűek ezeket az egyeztetéseket aligha fogják elrontani, ennek csak akkor van értelme, ha az egyeztetéssel sikerül további információkat

nyerni. Ez nyilván olyan szituációkban fordulhat elő, amikor több alanyesetű (esetrag nélküli) mondatrész van a mondatban és eldöntendő, hogy mi az alany. Van olyan szerencsés szóhasználat, amikor eldönthető, hogy mi micsoda; máskor viszont nem.

**Pl. Elvesztettem a kalapom.** (A "kalapom" tárgy, mert van birtokjele és nem egyezik a grammatikai alannyal.)

**Magad is látod.** (Eldönthetetlen a "magad" lehet alany is tárgy is.)

**Magad is látszol.** (A "magad" nem lehet tárgy mert tárgyas ragozást vonzana.)

**Péter kenyerét eszi.** (Eldönthetetlen, hogy Péter eszi-e, vagy Péterét.)

**Péter kenyerét eszik.** (Egyértelműen az utóbbi és az is látható, hogy az ík nem e.sz.3.sz., hanem t.sz.3.sz..)

A fentiekből látható, hogy az egyeztetés elvégzésére néha szükségünk van a kétértelműségek feloldásához. Viszont tapasztalatunk szerint a szöveg 80-90%-a nélkül is egyértelmű. Elképzelésünk szerint vagy úgy kell implementálni, hogy az egyértelmű esetekben rendkívül gyors legyen, vagy csak a kétértelmű esetekben fogjuk ezt a vizsgálatot elvégezni.

#### A birtokos szerkezet egyeztetése

A nem hiányos birtokos szerkezet birtokosból és birtokból, áll és a birtok birtokjelet visel. Ez megadja a birtokos személyét és számát és a birtok számát is. Ha a birtokos első vagy második személyű, a birtokrag teljesen meghatározza, és így nem szükséges kirakni. Ha a birtokos ki van téve, egyeztetni kell a birtokjellel. A birtok száma alapján megkülönböztetjük a következő eseteket:

egyes számú birtokos - egyes számú birtok  
egyes számú birtokos - többes számú birtok  
többes számú birtokos - egyes számú birtok  
többes számú birtokos - többes számú birtok.

Sajnos többes számú birtokos és egyes számú birtok esetén nem egyértelmű, hogy hány dologról van szó. Pl. **Elkérték a könyvünket.** (Lehet, hogy egy közzset, lehet, hogy mindenkítől egyet-egyet.) Ha ilyen kifejezés az alany, az állítmány egyes számú lesz. Viszont később többes számú hivatkozás is lehet rá (ha a második értelmezésről van szó). **Bemenetkor elkérték a könyvüket, és kilépéskor lepecsételve kapták vissza azokat.** Itt is megkülönböztetendő tehát a grammatikai és logikai szám, az utóbbi ilyen esetben bizonytalan.

Ha a birtokos számneves főnév, csoportos jelentésű vagy felsorolás, a birtokjelet a birtokos grammatikai számával egyeztetjük. Ilyen esetben, noha a birtokjel egy birtokost és egy birtokot mutat, mégis lehet logikailag többes számú. Pl. **Beszédte a három fiú könyvét, és azok elvesztek.**

Ha a birtokos felsorolás, amely több különböző személyből áll, az utolsóval egyeztetünk. Pl *Csak a te vagy a tanár kérésére engedem el.*

(Megjegyzések:

- A t.sz.3.sz.-ben a birtokos szerepében nem az *ők* hanem az *ő* személyes névmás áll. Pl. *az ő könyvük.*

- A *maga* névmás e.sz.3.sz.-jű birtokjelet követel. Pl. *a magam dolga.)*

A birtokos szerkezet egyeztetésére különösen az egymástól elszakadt birtokos és birtok összekapcsolásánál lehet szükség.

## A vonzatok

A szintaxis-leírás igei és névszói szerkezetekkel foglalkozó része szóosztályokon operálva általános sémákban írta le a mondat szerkezetét. Vannak ezzel szemben a helyes mondatnak olyan szabályai is, amelyeket nem annyira az általános szabályok határoznak meg, hanem a benne szereplő egyes szavak. Például az *ad* ige lehetséges argumentumai: hogy mit ad, kinek ad, miért cserébe, milyen okból vagy milyen célból. A *megy* ige lehetséges argumentumai ezzel szemben: honnan, hova megy, milyen útvonalon, milyen okból vagy milyen célból. Az argumentumok egy részét az jelöli ki, hogy az illető argumentum valamilyen ragot vagy névutót vonz. Az igének ezek a vonzatai annyira kötődnek az igéhez, hogy ugyanezek lesznek a vonzatai az igeneveknek és az igéből képzett más szavaknak is. Nemcsak az igéknek hanem más szavaknak, például melléknemeknek is lehet vonzata. Például, a *hasonló* szónál meg kell mondani, hogy mihez, a *generációs* szónál meg kell mondani hanyadik, a *hivatott* szónál, hogy mire.

A vonzatok egy sereg problémát vetnek fel. Az első és legfontosabb, hogy mi tarozik a vonzatok közé és mi nem. Vonzata-e az igének az alany, a tárgy és az összes határozó. A mi leírásunkban nem. Azokat a mondatrészeket, amelyek minden igéhez hozzákapcsolhatók nem tekintjük vonzatnak. (És hasonlóan a melléknévnek sem vonzata az *alig*, *kissé*, *majdnem* és hasonló szavak, mert ezek minden melléknévhez hozzá kapcsolhatók.) Mi csak azokat a szintaktikus egységeket tekintjük vonzatnak, amelyek arra az egyedi szóra jellemzők. Így hát nem vonzat az alany, a hely és az idő, ahol az ige végbemegy, a mód-, cél-, eredet-, eredmény- stb határozók, amelyek minden igéhez hozzákapcsolhatók.

Mi a vonzatokat egy szótárban kívánjuk tárolni. Amikor a szintaktikai elemzés során megállapítottuk, hogy az adott szintaktikai szerkezetbe valamilyen konkrét szó került és ennek vonzatai lehetnek, akkor ehhez a szótárhoz fordulunk és ennek alapján keressük meg a kapcsolódó komponenseket.

Természetesen nem ez az egyetlen lehetséges megközelítése a szintaxis leírásának. Elképzelhető lenne egy olyan leírás is ahol a szavak olyan részletesen lennének osztályozva, hogy egy osztályba csak az éppen azonos vonzatokkal rendelkező szavak kerülnének. Ekkor a szintaxis-leírásba konkrétan bele lehetne írni a vonzatokat. Ezzel szemben a szavak felosztása igen sok csoportra történe, és egy-egy szó igen sok csoportba beletartozna. (Mert ugyanannak a szónak több különböző vonzat-profilja is lehet.) Ezzel együtt a szintaxis-leírás is jelentősen megnagyobbodna.

A másik véglet az lenne, hogy a szótárban minden egyes szónál megfelelő részletességgel le lenne írva, hogy milyen más szavakhoz kapcsolódhat. Ez esetben külön szintaxis-leírásra nem lenne szükség.

Tisztán áttekinthetőség, kezelhetőség és hasonló gyakorlati szempontok alapján választottuk ezt a leírást. Úgy véljük megadható a mondatoknak egy általános leírása, és ettől elválasztva az adott témakör szavai, kiegészítve azokkal az információkkal, hogy az adott téma esetén, hogyan lehet ezeket a szavakat felhasználni (azaz milyenek a vonzataik).

A következő kérdés, hogy milyen vonzatai lehetnek egy szónak. A igék esetében a vonzat leggyakrabban egy ragos vagy névutós névszói csoport. Gyakori az olyan ige is amelyik egy másik igtét vonz, főnévi igenévi alakban. Pl. *Elfelejtett bevásárolni. Megtanult traktort vezetni.* Továbbá lehet a vonzat teljes mellékmondat is. Pl. *Elhiszi, hogy ...* . Melléknevek esetén lehet a vonzat számnév, pl. *hány méteres*, sorszámnev, pl. *hanyad osztályú*, melléknév, pl. *milyen minőségű*, tulajdonnev, pl. *milyen nevű* stb. De itt is gyakori a ragos névszói kifejezés pl. *mire képes*.

Kezelhetőségi szempontok alapján érdemes bizonyos vonzatokat egyetlen vonzat-osztállyá összevonni, például egyetlen osztálynak venni a honnan kérdésre válaszoló *-BŐL*, *-TŐL*, *-RŐL* ragos vagy a *mögül*, *alól*, *felől*, stb. névutós névszói kifejezéseket. (Természetesen ideértve a *tőle*, *rőla*, *mögülem* stb. alakokat is.)

Ezek szerint a vonzat lehet: egy konkrét szó, egy szóosztály, egy szóosztály valamilyen konkrét végződéssel vagy névutóval, főnévi igenév, mellékmondat, vonzat-osztály. (A vonzatrendeszer leírása körülbelül azonos azzal, amit az Apreszján-Páll féle orosz-magyar igei vonzatszótár alakalmaz.)

(Megjegyzés: Bizonyos esetekben pl. a között igekötővel kijelölt argumentumnál szükséges, hogy az logikailag többes számú legyen. Ilyenkor az argumentum vagy többes számú, vagy felsorolás, vagy több argumentum kötőszóval összekapcsolva.)

### Argumentum és vonzat

Mint azt már korábban megállapítottuk, az igéknek argumentumaik vannak, amelyek bizonyos szemantikus szerepeket töltenek be. A szintaxis-analízis egyik fő feladata, hogy ezeket az argumentumokat megtalálja. Az argumentumok felismerése a vonzatokon keresztül történik. Az adott argumentumhoz adott vonzat(ok) tartozanak és általában ezek jelölik ki az argumentumot, de sajnos az argumentum és vonzat között nincs egyértelmű kapcsolat. Egyrészt ugyanazt az argumentumot több különböző vonzattal is meg lehet adni. Például: *A folyadékokat literrel mérjük.* vagy *A folyadékokat literben mérjük.* Másrészt ugyanaz a vonzat több különböző argumentumhoz is tartozhat, pl. *Gyomorfekéllyel kezelte magát.* *A híres belgyógyással kezelte magát.*



Tovább nehezíti a helyzetet, hogy például a **-NAK** rag nemcsak igei vonzatként léphet fel, hanem az elváló birtokos ragjaként is, és bizonyos segédigék esetében az alany is **-NAK** ragot kap.

A melléknevek és melléknévi igenevek esetén a vonzat mindig a szó előtt áll, de ha több vonzat is van (ami a gyakorlatban igen ritkán fordul elő) ezek sorrendje általában nem meghatározott. Pl. *a főnököket beadványaival ostromló beosztott, a beadványaival a főnököket ostromló beosztott.*

Amikor a szintaxis-elemzés során eljutunk egy olyan pontra, hogy egy adott ige argumentumait össze kell szednünk, több olyan szószerkezetet is találhatunk, ami potenciálisan argumentum lehet. Ekkor négy eset lehetséges: vagy több argumentumunk van (felsorolás vagy kötőszavas szerekezet formájában), pl. *Pista vagy Jancsit vagy Jóskát látta a piacon.*, vagy egy argumentumunk van hátravetett jelzővel, pl. *Pistát, régi barátját látogatta meg.*, vagy valamelyik argumentum, a másik pedig nem, pl. *Pistának kellett enni adnia a disznóknak.*, vagy egyik sem argumentum, pl. *Pistának a levelét nekem kellett elvinnem.* (Feltéve, hogy Pista írta és én viszem.)

Az ilyen mondatok szintaktikailag kétértelműek. A kétértelműségek kiküszöbölésére a leghatékonyabb és legbiztosabb út, a szemantikus analízis, amely megállapíthatja, hogy bizonyos mondatfelépítések értelmezhetők mások viszont nem. Például, ha tudjuk, hogy az *ad* ige egyik lehetséges vonzatformája: *ad+személy+élőlény-NEK+(enni/inni)*, akkor a *Pistának a disznóknak kellett enni adnia.* mondatban egyértelműen Pista lesz az ágens és a *disznók* a dativusz.

A fenti szemantikai elemzéshez ún. szemantikai jegyek bevezetésére van szükség. Noha bevezethetők lennének olyan általános szemantikai jegyek, amelyek általában témakörből és nyelvtől függetlenek, mi a gyorsabb és pontosabb elemzés érdekében erősen a témakörhöz kötött jegyek alkalmazására gondolunk. (A szemantikai jegyekre egy későbbi fejezetben visszatérünk.)

A másik kérdés, hogy a fent leírt analízist mikor végezzük; először összeállítjuk-e az összes lehetséges fát, és azután végezzük-e el az analízist, vagy már a vonzatok keresése közben tekintettel vagyunk a szemantikai jegyekre. Hatékonysági megfontolások alapján az utóbbi megoldás látszik célszerűnek.

A kétértelműségek feloldásában a szemantikus jegyeken kívül más ad hoc szabályok alkalmazása is szóba jöhet, például azé a szabályé, hogy a birtokjeles szó előtt álló **-NAK** ragos névszó nagy valószínűséggel birtokos lesz. Ilyen és hasonló szabályok alkalmazása azért problematikus, mert a magyar nyelv szabályait nem egy programmal akarjuk definiálni, hanem egy géptől és programnyelvtől független formalizmussal, amelybe az ad hoc megoldások nehezen illeszthetők bele.

### Kötelező és opcionális argumentumok, defaultok

A továbbiakban vonzat-profilnak fogjuk nevezni a vonzatoknak egy olyan maximális készletét, ami egy mondaton belül lehetséges, de további elemekkel nem bővíthető. Egy szónak több vonzat-profilja is lehet. Két vonzat-profil akkor különbözik egymástól, ha az egyikben van olyan vonzat ami a másikban nincs, vagy ha ugyanaz az a vonzat más szerepet játszik az egyikben mint a másikban. Pl. *valami valamire szolgál, valaki valakit szolgál; összehasonlít valamit valamivel, összehasonlít (több) valamit; feltölt valamit valamilyen anyaggal (vízzel töltötte fel a medencét), feltölt valamit valamilyen eszközzel (vödörrel töltötte fel a medencét).*

A vonzat-profilon belül vannak kötelező és opcionális argumentumok. Kötelező az az argumentum, amelyik szükséges a mondatok értelmezéséhez, és opcionális, amelyekre nincs feltétlenül szükségünk. (Mint látható, a fenti definíció igen szubjektív és nagyon függ attól, milyen adatkészleten fogunk dolgozni.) Ha egy opcionális argumentum hiányzik, helyére egy előre meghatározott default elemet teszünk. Van egy univerzális elemünk; a meghatározatlan az esetek jelentős részében ez lesz a default, de nem mindig. Pl. *Pista elmegy moziba. elmegy+honnan+hova+mivel* és mindhárom argumentuma opcionális. A *mivel* argumentum meghatározatlan lesz, a *honnan* ezzel szemben az alany aktuális helye lesz.

Ha ezzel szemben egy kötelező argumentum hiányzik, akkor ennek helyébe egy speciális jelzés kerül, ami arra utal, hogy a mondat hiányos, és ez a hiány egy implicit hivatkozás valamire, amit meg kell találnunk. (Ezzel a kérdéssel a hivatkozásokról szóló fejezet foglalkozik.)

(Megjegyzések:

Kötelező argumentum lehet olyan "Hol?" kérdésre válaszoló helyhatározó vagy időhatározó is, amit más esetekben nem tekintünk argumentumnak. Pl. *valami valahol található.*

Az ige összes argumentuma az igeneveknek és az igéből képzett *-ás,-és* végű főnévnek is argumentuma, sőt a hely és időhatározó is argumentuma lesz ezeknek, de egyik argumentum sem kötelező.)

## A szintaxis analízis megvalósítása

Mint azt az előző fejezetekből láttuk, a magyar nyelv szintaxisának két szintje van. Az alsóbb szinten a névszói szerkezetben a pozicionális meghatározottság majdnem teljes, ezzel szemben a felsőbb szinten az összetartozás nagyrészt ragokkal (és névutókkal) van definiálva. Ez indokolja azt, hogy a két szint analízisét elkülönítsük, és két különböző módszerrel oldjuk meg. Mindkét feladat során fellép viszont egy közös probléma: a vonzatok megtalálása. A kötött szórendű részben a vonzatok a vonzattal rendelkező szó elé kerülnek, a szabad szórendű részben viszont végződésük alapján kell a szóhoz kapcsolni őket.

Elképzelésünk szerint a szintaxisfa felépítése két menetben történne: az egyik menetben összeszednénk pozicionális alapon a névszói kifejezések részeit, a következő menetben pedig felülről az ige felől elindulva, összeszednénk az igei kifejezést, az ige vonzatait és a szabad határozókat.

A névszói kifejezés komponenseinek összeszedése egy ATN, vagy valami ehhez hasonló elemzőeszköz segítségével történhet.

(Az ATN (Augmented Transition Network) mind a természetes nyelvek, mind a programozási nyelvek szintaxisának jólismert leíró/elemező eszköze. Az elemek sorrendjét egy végesautomata-szerű sémával írjuk le, ahol azonban az egyes állapotok közötti átmenet vagy egy szimbólum hatására történik vagy az átmenetet magát is egy hasonló automatával írhatjuk le. Ezenkívül arra is lehetőséget nyújt, hogy globális változókon keresztül információkat továbbítsunk.)

Mivel a magyar nyelvben a pozicionálisan kötött vonzatok mindig a lexikai elemtől balra állnak, célszerűnek látszik, ha az elemzés a szokásostól eltérőleg jobbról balra halad.

Ezek szerint a névszói kifejezések megkeresése vázlatosan a következő lépésekben történik. Beolvassuk a mondatot és elvégezzük a morfológiai analízist. Ezek után, a legutolsó esetragnál vagy névutóból indulunk ki. Ez előtt egy főnév állhat, az előtt egy anyag-, nemzet- vagy foglalkozásnév, az előtt egy tulajdonnév, az előtt tulajdonságjelzők, az előtt számjelző, az előtt névelő. Ha valamelyik elmarad, csak az előtte álló következhet. Lehetnek azonban a sorban vonzattal járó szavak is. Ilyenek az igéből képzett főnév (pl. *iskolába járás* a melléknévi igenevek, a melléknévi névutó (pl. *a ház előtti fa*), vonzattal járó melléknév (pl. *eredményeire büszke vezető*). A vonzat lehet kötelező vagy opcionális. (Pl. *a méretű* melléknév vonzata kötelező, a *büszke* szóé opcionális.) A vonzat előírja, hogy milyen kategóriájú szó vagy kifejezés állhat az adott szó előtt. Az esetek nagy részében a vonzat maga is egy ragos névszói kifejezés. Az ilyen beágyazott szerkezetek elemzése után vissza kell térnünk a magasabb szintű szerkezet elemzésére. Amikor a teljes névszói kifejezést megtaláltuk, ugyanezt az eljárást megismételve további névszói

kifejezéseket keresünk a mondat megmaradt elején.

Ez a módszer két technikai problémát vet fel. Az első az, hogy az alanyesetnek nincs ragja. Ezért az eseterag nélkül álló soronkövetkező névszót úgy kell kezelniük, mint egy potenciális névszói kifejezés végét.

A második probléma, hogy ha a névszói kifejezés egy opcionális argumentummal kezdődik, akkor ez vajon az utána álló vonzatos szóhoz tartozik-e, vagy esetleg egy magasabb szintű konstrukcióhoz például az ígéhez. Pl. *Az utcán játszadozó gyerekeket látok.* (Én vagyok az utcán, vagy a gyerekek, vagy mindannyian ott vagyunk.) Ehhez hasonlóan meg kell vizsgálni, hogy ha a soronkövetkező szó többértelmű, az adott pozíció meghatározza-e egyértelműen, hogy az adott esetben melyik értelme jön szóba. A analízisnek biztosítania kell, hogy különböző összevonások jöthessenek létre és a több lehetséges értelmezés megmaradjon.

Ezek után a analízis úgy mehet végbe, hogy megállapítjuk, milyen igei konstrukció szerepel a mondatban (ragozott ige, ragozott segédige, nem ragozott segédige, explicit VAN ige, ki nem tett VAN ige stb.). Majd megállapítjuk az ige vagy az egyéb vonzattal rendelkező szó vonzatait, és megvizsgáljuk, hogy a többi kifejezés szabad határozónak tekinthető-e.

(Megjegyzés: ez a szintaxis-leírás nem foglalkozik a határozói igenévvel általában és általában a szabad módhatározókkal (állapot-, ok-, cél-, eredmény-, stb.). Ennek az az oka, hogy ezek szemantikus értelmezése problémát jelent.)

A program mindenesetre három komponensből fog állni: egy ATN-szerű pozícionális szabályokat kezelő eszközből, egy vonzatokat kezelő eszközből, és egy a szabad sorrendű szabályokat kezelő eszközből.

Felmerül a kérdés: meg lehet-e a teljes elemzést valamilyen ismert és egységes eszközzel oldani. Szóba jöhet például valamilyen ismert CF- elemző. Ezekkel az a probléma, hogy a programozási nyelvekre létrehozott elemzők a balról jobbra egyértelműen elemezhető nyelvekre készültek. A LISP nyelv alkalmazása csak a programozás stílusát változtatná meg ahhoz képest, mintha egy Pascal vagy MODULA nyelvben implementált listakezelőt használnánk. Felmerül a lehetőség a PROLOG nyelv alkalmazására, meg kell azonban vizsgálnunk, hogyan írhatók le a szabad szórendet kifejező szabályok. Elképzelhető még egy Production System jellegű rendszer alkalmazása is. Az utóbbi kettőnél a nyelv szintaktikai szabályai alkotnák a program egyik fő komponensét. Szándékunkban áll ilyenekkel kísérleteket végezni, de tartunk tőle, hogy a hatékonyság nem lesz megfelelő.



Az utóbbi kettőnél az a fő probléma (amely előny a kísérleti megvalósításoknál), hogy a programozó csak azt írhatja elő, hogy milyen összefüggéseket kell megállapítani, és nem befolyásolhatja, vagy csak igen nehezen, hogy milyen stratégiával, milyen sorrendben kell az összefüggéseket megkeresni. A minden utat végigjáró és a zsákutcákból visszalépő algoritmusok igen általánosak és kevésbé hatékonyak. A hatékonyság érdekében viszont arra kell törekednünk, hogy amit csak lehet, egyetlen menetben állapítsunk meg.

Ha a szintaktikai elemzés valamilyen tradicionális nyelven valósul meg, gondoskodni kell a szintaxis-leírás könnyű gyors változtatásáról, javíthatóságáról. Olyan megoldást kell választani, hogy vagy az elemző program adatként kapja meg a szintaxis-leírást, vagy (és ez a valószínűbb) az elemzőprogramot egy generátor-program állítsa elő a szintaktikus leírásból.



A természetes nyelv egyik ismert tulajdonsága, hogy nem beszélünk kerek mondatokban, vagyis nem adunk meg minden információt, amely a mondat megértéséhez szükséges, hanem előzőleg megadott információkra hivatkozunk.

A hivatkozások feloldása igen bonyolult feladat, mert a hivatkozásoknak számos fajtája van és kezelésük rendkívül szerteágazó problémákat vet fel. A következőben összefoglaljuk a legfontosabb problémákat.

A hivatkozásoknak a magyar nyelvben is (más nyelvekhez hasonlóan) három alapvető formája van. A első forma a névmás, amely definíciója alapján egy névszói kifejezést pótol (vagy annak egy részét). A második forma a hiány, nyelvészeti szakkifejezéssel az ellipszis, amikor az adott mondatrész hiányzik a mondatból, de oda kell érteni. Pl. *Látod azt az embert? Látom.* A harmadik lehetséges forma a határozott névszói kifejezés. Ilyen esetben a határozott névszói kifejezés utal egy másik névszói kifejezésre, amely korábban már szerepelt a szövegben. Pl. *Van-e folyó New York-ban? Van. Mi a neve a folyónak?*

A harmadik fajta (a határozott névszói) hivatkozás a legproblematisabb mert, mert nem minden határozott névszói kifejezés hivatkozás. Pl. *Az ember halandó.* mondatban az ember nem utal semmire. Sajnos egyelőre semmilyen leprogramozható kritériumot nem tudunk mondani arra, hogy mikor visszautaló egy névszói kifejezés.

A visszautalás természetesen nemcsak ugyanazzal a szóval történhet mint amit a korábbiakban használtunk, hanem szinonimákat is használhatunk, sőt mindenféle más azonos értelmű kifejezést is. Pl. *Pista és Jóska bement a moziba. A két gyerek a 10. sorban ült le.* A visszautaló névszói kifejezés gyakran maga is hiányos. Pl. *Melyik nyakkendőt vegyem meg? A kékét!*

Ha a számítógéppel dialógust folytatunk, természetesen nemcsak azokra a dolgokra hivatkozhatunk vissza, amit mi kérdeztünk, hanem azokra a dolgokra is, amit a gép felelt.

Ezen túlmenően, nemcsak olyan dolgokra hivatkozhatunk, ami explicite előfordult a korábbi szövegben, hanem hivatkozhatunk egy egész mondatra egyetlen szóval, sőt olyan dolgokra is, amelyekről egyáltalán nem volt szó, de a korábbiakból kikövetkeztethetők. Pl. *Pista repülővel utazott Amerikába.*  
- *Milyen hosszú volt az utazás? Mennyibe került a jegy?*

Vannak ugyan olyan természetes nyelvű interfészek, amelyek ezeket a visszahivatkozásokat egyáltalán nem kezelik, de ezek tulajdonképpen igen távol vannak a természetes nyelvtől. Manapság egy interfész tudását nagyrészt aszerint ítélik meg, hogy milyen szabadságot nyújt ezen a téren.

Noha a probléma teljes általánosságban való kezelése rendkívül nagy apparátust igényel, bizonyos eredmények már viszonylag egyszerű eszlőzőkkel is elérhetők. Ezek közé tartoznak a nyelvtani szabályok és a szemantikai egyeztetés.

Az európai nyelvekben a leggyakrabban alkalmazott eszköz a nemben, és számban történő egyeztetés. Ez sajnos a magyarban egyáltalán nem alkalmazható, mert a magyarban nincs nem, a számbeli egyeztetés ingadozik az ún. logikai és grammatikai egyeztetés között. (Mondaton belül az egyeztetés inkább grammatikai, mondatok között általában logikai szám szerint egyeztetünk. Lásd "Egyeztetések".) Hasonlóan, élettelen dolgokra is gyakran hivatkozunk ő névmással és élőkre az-zal.

A nyelvtani szabályok közül a következőkre lehetünk tekintettel:

Ha úgy képzeljük, hogy a mondat egy ige szerkezetből és annak bővítményeiből áll (alanyból, tárgyból, határozókbókból), ezek hivatkozhatnak egymásra és a következő mondat is hivatkozhat rájuk. Az ezekbe beágyazott igei szerkezetek (almondatok és igeneves szerkezetek) bővítményei hivatkozhatnak, a főszerkezet bővítményeire, és saját társ bővítményeikre, de nem hivatkozhatnak egy másik beágyazott szerkezet belső elemére. A következő mondat sem hivatkozhat az előző mondatba beágyazott szerkezet belsejére.

Ez alól némileg kivételt képez a főlső szintű bővímény birtokosa, amire mégis szokás hivatkozni.

Arra nézve, hogy egy visszahivatkozásnál mikor használunk hiányt, mikor személyes- és mikor mutatónévmást, valamint ezekhez hasonló dolgokra É. Kiss Katalinnál és Pléh Csabánál számos hipotézist találunk. (Pl. Ha két egymást követő mondat alanya azonos, akkor a második mondatban nem tesszük ki. Ha a második mondat alanya azonos valamivel ami az előző mondatban nem alany, akkor az az névmással hivatkozunk rá.) Nekünk azonban úgy tűnik, hogy ezek a szabályok inkább tendencia jellegűek, és csak szemantikus egyeztetéssel együtt alkalmazhatók.

A szemantikai egyeztetés lényege, hogy minden névszói kifejezés bizonyos szemantikai jegyeket hordoz, az a hely pedig, ahol hivatkoznak rá, bizonyos szemantikai jegyeket kíván meg. A hivatkozottat tehát azok között kell keresni, akik a szükséges szemantikai jegyeket hordozzák.

Fontos megfigyelés még, hogy az említett dolgokra, ha újra nem került rájuk szó, az idő előrehaladtával egyre kevésbé lehet hivatkozni, bár a "felejtés" gyorsasága emberenként változó.

Ezek alapján az első programváltozatban a hivatkozások feloldását a következő lépésekben képzeljük el: amikor a számítógép kap egy mondatot, a főmondat bővítményeit feltesszük egy hivatkozási listára. Az eredményül kapott objektumokat szintén erre a listára tesszük. A következő mondatban szereplő objektumokat szintén erre a listára tesszük. Ha a mondatban hivatkozás van, akkor ezek közül keresünk olyat, ami szemantikailag megfelelő jegyeket hordoz. Ha több ilyen találunk, vagy egyet se, visszakérdezzük a felhasználótól. Ha egyet találunk, azt tekintjük a hivatkozottnak. A következő kérdés előtt a kettővel korábbi kérdést és a rá adott választ töröljük a hivatkozható dolgok listájáról.

## Szemantikus reprezentáció és szemantikus jegyek

A szemantikáról igen nehéz beszélni a konkrét témakör ismerete nélkül. Elképzelhető ugyan egy olyan szemantikus definíció, amely szélességében (sok témakörre kiterjedően) eléggé jól átfogja a teljes magyar nyelvet, de ez a definíció nagyon az általánosságok szintjén fog mozogni, azaz mélységében igen sekély lesz. Nyilvánvaló, hogy a természetes nyelv egész mást jelent egy matematikusnak, nyelvésznek vagy jogásznak amikor saját szakterületéről van szó (ilyenkor talán még sem értik azt, amiről a másik beszél), és mást, amikor egy közös témaköréről, például egy külföldi utazásról van szó.

Mint azt már korábban többször lerögzítettük, mi a programunkban egy témakör- és feladat-specifikus szemantikára gondolunk, és mivel a feladat jelenleg nincs még kiválasztva, nagyon nehezen tudunk róla beszélni. Az alábbiakban példákon keresztül fogjuk megvilágítani, mire is gondolunk, amikor szemantikáról beszélünk.

Tegyük fel, hogy van egy geometriai alapú rajzoló programunk, amelyet természetes nyelvű utasításokkal próbálunk vezérelni. Olyan parancsokat képzelhetünk el mint: *Kösd össze a háromszög AB oldalának felezőpontját a szemben lévő csúccsal!*, *Hosszabbítsd meg az AB szakaszt a három és félszeresére!*, *Rajzolj egy szaggatott felező vonalat az AB szakasz közepére!* stb.

Nyilvánvaló, hogy mind a három fenti parancs a rajzoló programnak ahhoz a lehetőségéhez kapcsolódik, hogy húzz vonalat valahonnan, valameddig. (Azt, hogy milyen legyen ez a vonal: szaggatott, kék, vastag, stb. az rendszerint egy másik utasítása a rajzoló programnak. Hasonlóan, a felezőpont, a szemben lévő csúcs megkeresése, sőt esetleg az A és a B pont koordinátáinak megkeresése, stb. külön tennivaló, amelyre vagy a rajzoló programot kell utasítani, vagy az interfésznek kell megoldania.) Mind a három mondatot vissza kell tehát vezetnünk egy olyan sémára amelyben adva van egy parancs, hogy "rajzolj" és három argumentum: mit (egyenesszakaszt, körívet, illeszkedő görbét, stb.), mitől, meddig. Az argumentumok meg lehetnek adva közvetlenül felhasználható operandusokkal, vagy újabb parancsok sémáival, amelyek végrehajtása a szükséges argumentumot szolgáltatni fogják.

A szemantikus reprezentáció elkészítése során, azt is ellenőriznünk kell, hogy a megadott parancs értelmes és teljes-e. Nem fogadhatunk el olyan parancsokat mint: *\* Húzzál háromszöget az A és B pont közé!*, *\* Kösd össze az AB szakasz felezőpontját piros ponttal!*, *\* Felezd meg a háromszöget!* stb. Ennek az ellenőrzésnek két oka is lehet: vagy arról van szó, hogy a rajzoló program bemenete primitív (pl. csak abszolút koordinátákat tud összekötni), ekkor nekünk magunknak kell erre a primitív nyelvre a mondatot lefordítanunk, és ez hibás parancs esetén nem sikerülhet; vagy a rajzoló program



bemeneti nyelve bonyolult dolgokra is lehetőséget ad, és látszólag át tudnánk írni rá a hibás parancsot is, de a rossz parancs rossz működést váltana ki, és ennek következményeit kellene viselnünk (pl. a meghívott program abortál) .

A parancs akkor értelmes, ha jók az argumentumai. Ellenőriznünk kell, hogy minden argumentum megvan-e, és olyan-e, mint amilyennek lennie kell. Esetünkben vannak síkidomok, vonalak, pontok, színek, vonalstílusuk (folytonos, szaggatott, pontozott, stb.), vonalvastagságok, távolságok, stb. Ellenőrizni kell, hogy ahova pont szükséges, oda pont kerüljön, ahova szín, oda szín, ahova távolság, oda távolság.

Éppen ezért a felmerülő fogalmakat osztályokba kell sorolni. Az osztályba sorolás hierarchikus is lehet, például: geometriai fogalmak, azon belül síkidom, vonal, pont; a síkidomokon belül négyszög, háromszög, kör; a négyszögon belül paralellogramma, téglalap, négyzet, stb. Azt, hogy egy fogalom milyen osztályba illetve osztályokba tartozik, szemantikus jegyeknek nevezzük.

A parancsokat leíró sémák előírják, hogy milyen argumentum pozícióba milyen szemantikus jegyeket viselő argumentum szükséges. Például a rajzolj parancs mit argumentuma csak valamilyen vonal lehet, a másik két argumentuma csak pont.

Természetesen egy másik témakör esetén mások lesznek a parancs-sémák és mások lesznek a szemantikus jegyek. Például egy térképészeti adatbázis esetén a következő fogalmakkal dolgozhatunk: ország, város, főváros, hegy, víz, folyó, tó, terület, hosszúság, lakosság. Megkérdezhetjük, hogy egy tónak mekkora a területe, de nem kérdezhetjük meg, mekkora a lakossága. Egy folyóról megkérdezhetjük, mekkora a hossza, de nem kérdezhetjük meg, mekkora a területe. stb.

Felmerül a kérdés, hogy a szemantika leírására szolgáló sémákat milyen formában ábrázoljuk a számítógépben. Több megoldás képzelhető el, a logikai formuláktól kezdve, a fogalmi hálókig. Mi pillanatnyilag frame-ekben képzeljük el a megvalósítást.

Ezt az ábrázolást azután le kell fordítanunk annak a programnak az input nyelvére, amelyhez az interfészt készítettük. Egy mondatból az esetek nagy részében több utasítás is lesz, ezeket az utasításokat olyan sorrendben kell kiadni, hogy a későbbi a korábbi eredményét fel tudja használni. Ezt a fordítást egy mechanikus tevékenységnek képzeljük el, melynek mérete azért nem elhanyagolható.



# Melléklet a kísérleti morfológiai elemző toldalékairól.

A melléklet három táblázatot tartalmaz. Az első az általunk kezelt töváltozásokat, a második a tövégi betűhalmazokat, a harmadik pedig magukat a toldalékokat sorolja fel toldalékosztályonként. A toldalékok előtti szám a toldalékot megelőző betűhalmazát (esetleg betűhalmazok metszetét) jelöli, míg az azt követő szám(ok) a lehetséges töváltozás(oka)t jelöli(k).

## Töváltozások (inverz):

- 1 igei hasonulás zz ss ssz ddz ggy után, + igei s-t (opcionális) (fut de tát fos)
- 2 névszói áa ée ill +a +e (opcionális mert lé, -é, de á-nál kötelező)
- 3 névszói hasonulás dupla mássalhangzó után (ha nem talál olyan tövet opcionális)
- 4 névszói aó eő +ó +ő (ajtó apró tető erdő)
- 5 igei v előtt oó őő üü uú úú íí íí +sz (ró lő nyű fú ri alsz esz isz)
- 6 névszói v előtt oó aó őő őő uú üü üü eé +u +ü (ló tó mű bő tő fű szű lé falu tetű)
- 7 ú ü kiesés (opcionális) (borjú fiú lassú hosszú szörnyű) + névszói belső eé aá uú üü íí (tér nyár úr tűz híd)
- 8 o e ő kiesés (opcionális) jlrms (irodalom forog késedelem képez köröm sör)

## Betűhalmazok:

- |        |          |        |                    |                    |                      |
|--------|----------|--------|--------------------|--------------------|----------------------|
| 0 -aá  | bc deé   | fghií  | jklmn              | oóőőpqrstuúüüvwxyz |                      |
| A Á BC | DEÉFGHIÍ | JKLMN  | OÓŐOPQRSTUÚÜÜVWXYZ | 0123456789         | {bármí}              |
| 1 -aá  | bc deé   | fghií  | jklmn              | oóőőpqrstuúüüvwxyz |                      |
| A      | E        |        | Ó Ő                |                    | {magyar vég}         |
| 2 -    | bc d     | fg h   | jklmn              | pqrst              | vwxyz{mássalhangzók} |
| 3 - á  |          | é      | hi                 | j l n oó ő         | rs uúüüv yz          |
|        |          |        | Ó Ő                |                    | {félmagánhangzók}    |
| 4 -a   |          | e      | i                  | l o őő             | r tu üü              |
| A      |          |        |                    |                    | {v-s névszó előtt }  |
| 5      | b de     | fg hií | jklmn              | oóőőp rstuúüüv     | yz                   |
|        | E I      |        |                    |                    | {igető vég}          |
| 6      | b d      | fg h í | jklmn              | ó őp r             | ú üv y               |
|        |          |        |                    |                    | {-e-i-s-t-z}         |
| 7 -aá  |          | eé     | hií                | oóőő q             | uúüü y               |
| A      | E        |        | Ó Ő                |                    | {magánhangzók}       |
| 8 -aá  |          |        | íí jklmn           | oó                 | rs uú z              |
| A      |          |        | Ó                  |                    | {Mély magánhangzók}  |
| 9 -    |          | eé     | íí jklmn           | őő                 | rs üü z              |
|        | E        |        | Ő                  |                    | {Magas magánhangzók} |
| 10 - á | bc d é   | fg hií | jklmn              | oóőőpqrstuúüüvwxyz |                      |
|        |          |        | Ó Ő                |                    | {-a-e}               |

Toldalékok:

*birtokos	*névszónévszó	*névszóige	*igenévszó
0 0	0 i 8	1 ít 8 7	2*5 andó 8
10 é 2	0 as 8 7	2 az 8 7	2*5 atlan 8
10 éi 2	0 es 8 7	2 ez 8 7	2*5 ás 8
	0 os 8 7	2 oz 8 7	2*5 endő 8
	0 ős 8 7	2 őz 8 7	2*5 etlen 8
*igeige	2 atlan 8 7	7*10 z 2	2*5 ett 8
2*5 egat 8	2 etlen 8 7		2*5 és 8
2*5 ogat	2 talan 0		2*5 ott 8
2*5 öget 8	2 telen 0		2*5 ó 8
5 at 0	7*8*10 tlan 2		2*5 őtt 8
5 et 0	7*9*10 tlen 2		2*5 ő 8
5 gat	4 vas 6		2*5 t 0
5 get 0	4 vatlan 6		7*5 tt 0
5 hat 0	4 ves 6		8*5 vandó 5
5 het 0	4 vetlen 6		8*5 vatlan 5
5 tat 0	7*10 s 2		8*5 vás 5
5 tet 0			9*5 vendő 5
5*7*9 veget 5	*fokozó		9*5 vetlen 5
5*7*8 vogat	2 abb 8 7		7*5 vett 5
5*7*9 vöget 5	2 ebb 8 7		9*5 vés 5
	2 b 0		7*5 vott 5
	4 vabb 6		8*5 vó 5
	4 vebb 6		7*5 vött 5
	7*10 bb 2		9*5 vő 5
*eset			
0 0	10 ba 2	10 ban 2	10 be 2
10 ben 2	10 ből 2	10 ből 2	10 ért 2
10 hez 2	10 hoz 2	10 hőz 2	10 ig 2
10 ként 0	10 kor 0	10 nak 2	10 nál 2
10 nek 2	10 nél 2	10 ra 2	10 re 2
10 ról 2	10 ről 2	10 tól 2	10 től 2
2 _al 3	2 _á 3	2 _el 3	2 _é 3
2 an 8 7	2 at 8 7	2 en 8 7	2 et 8 7
2 on 8 7	2 ot 8 7	2 őn 8 7	2 öt 8 7
3 t 2	4 vat 6	4 vet 6	4 von 6
4 vön 6	7*10 n 2	7*10 vá 2	7*10 vé 2
7*8*10 val 2	7*9*10 vel 2		

*igerag			
5 0	5 ja 0	5 ják 0	5 játok 0
5 juk 0	5 jük 0	5 lak 0	5 lek 0
5 na 0	5 nak 0	5 ná 0	5 nád 0
5 nák 0	5 nál 0	5 nálak 0	5 nám 0
5 nának 0	5 nánk 0	5 nátok 0	5 ne 0
5 nek 0	5 né 0	5 néd 0	5 nék 0
5 nél 0	5 nélek 0	5 ném 0	5 nének 0
5 nénk 0	5 nétek 0	5 ni 0	5 nia 0
5 nie 0	5 niuk 0	5 niük 0	5 nod 0
5 nom 0	5 notok 0	5 nõd 0	5 nõm 0
5 nõtök 0	5 nunk 0	5 nünk 0	5 sz 0
5 tok 0	5 tők 0	5 va 0	5 ván 0
5 ve 0	5 vén 0	5*2 _ 1	5*2 _a 1
5*2 _ad 1	5*2 _ak 1	5*2 _alak 1	5*2 _am 1
5*2 _anak 1	5*2 _atok 1	5*2 _ák 1	5*2 _ál 1
5*2 _átok 1	5*2 _d 1	5*2 _e 1	5*2 _ed 1
5*2 _ek 1	5*2 _elek 1	5*2 _em 1	5*2 _en 1
5*2 _enek 1	5*2 _etek 1	5*2 _ék 1	5*2 _él 1
5*2 _étek 1	5*2 _on 1	5*2 _uk 1	5*2 _unk 1
5*2 _ük 1	5*2 _ünk 1	5*2 ana 8	5*2 anak 8
5*2 aná 8	5*2 anád 8	5*2 anák 8	5*2 anál 8
5*2 análak 8	5*2 anám 8	5*2 anának 8	5*2 anánk 8
5*2 anátok 8	5*2 anék 8	5*2 ani 8	5*2 ania 8
5*2 aniuk 8	5*2 anod 8	5*2 anom 8	5*2 anotok 8
5*2 anunk 8	5*2 asz 8	5*2 attalak 8	5*2 d 0
5*2 ed 8	5*2 ek 8	5*2 el 8	5*2 em 8
5*2 ene 8	5*2 ened 8	5*2 enek 8	5*2 enem 8
5*2 enetek 8	5*2 ené 8	5*2 enéd 8	5*2 enék 8
5*2 enél 8	5*2 enélek 8	5*2 eném 8	5*2 enének 8
5*2 enénk 8	5*2 enétek 8	5*2 eni 8	5*2 enie 8
5*2 eniük 8	5*2 enünk 8	5*2 esz 8	5*2 etek 8
5*2 ett 8	5*2 ette 8	5*2 etted 8	5*2 ettek 8
5*2 ettelek 8	5*2 ettem 8	5*2 ettetek 8	5*2 ették 8
5*2 ettél 8	5*2 ettétek 8	5*2 ettük 8	5*2 ettünk 8
5*2 i 8	5*2 ik 8	5*2 itek 8	5*2 ned 0
5*2 nem 0	5*2 netek 0	5*2 od 8	5*2 ok 8
5*2 ol 8	5*2 om 8	5*2 otok 8	5*2 ott 8
5*2 otta 8	5*2 ottad 8	5*2 ottak 8	5*2 ottalak 8
5*2 ottam 8	5*2 ottatok 8	5*2 ották 8	5*2 ottál 8
5*2 ottátok 8	5*2 ottuk 8	5*2 ottunk 8	5*2 õd 9
5*2 ök 8	5*2 õl 8	5*2 õm 8	5*2 õtök 8
5*2 õtt 8	5*2 õtte 8	5*2 õtted 8	5*2 õttek 8
5*2 õttelek 8	5*2 õttem 8	5*2 õttetek 8	5*2 õtték 8
5*2 õttél 8	5*2 õttétek 8	5*2 õttük 8	5*2 õttünk 8
5*2 t 0	5*2 ta 0	5*2 tad 0	5*2 tak 0
5*2 talak 0	5*2 tam 0	5*2 tatok 0	5*2 ták 0
5*2 tál 0	5*2 tátok 0	5*2 te 0	5*2 ted 0
5*2 tek 0	5*2 telek 0	5*2 tem 0	5*2 tetek 0
5*2 ték 0	5*2 tél 0	5*2 téték 0	5*2 tuk 0
5*2 tunk 0	5*2 tük 0	5*2 tünk 0	5*2 unk 8
5*2 ünk 8	5*6 j 0	5*6 jad 0	5*6 jak 0
5*6 jalak 0	5*6 jam 0	5*6 janak 0	5*6 játok 0
5*6 jál 0	5*6 je 0	5*6 jed 0	5*6 jek 0
5*6 jelek 0	5*6 jem 0	5*6 jen 0	5*6 jenek 0

5*6 jetek 0	5*6 jék 0	5*6 jél 0	5*6 jétek 0
5*6 jon 0	5*6 junk 0	5*6 jünk 0	5*7 tt 0
5*7*9 vi 5	5*7*9 vik 5	5*7*9 vitek 5	5*7*8 vod 5
5*7*8 vok 5	5*7*8 vom 5	5*7*9 vöd 5	5*7*9 vök 5
5*7*9 vöm 5	5*7*8 vunk 5	5*7*9 vünk 5	5*7*8 tta 0
5*7*8 ttad 0	5*7*8 ttak 0	5*7*8 ttalak 0	5*7*8 ttam 0
5*7*8 ttatok 0	5*7*8 tták 0	5*7*8 ttál 0	5*7*8 ttátok 0
5*7*8 ttuk 0	5*7*8 ttunk 0	5*7*9 tte 0	5*7*9 tted 0
5*7*9 ttek 0	5*7*9 ttelek 0	5*7*9 ttem 0	5*7*9 ttetek 0
5*7*9 tték 0	5*7*9 ttél 0	5*7*9 ttétek 0	5*7*9 ttük 0
5*7*9 ttünk 0			

\*többses

0 0	1 ja 2 4	1 jai 2 4	1 jaid 2 4
1 jaik 2 4	1 jaim 2 4	1 jaink 2 4	1 jaitok 2 4
1 je 2 4	1 jei 2 4	1 jeid 2 4	1 jeik 2 4
1 jeim 2 4	1 jeink 2 4	1 jeitek 2 4	1 juk 2 4
1 jük 2 4	2 a 8 7	2 ad 8 7	2 ai 8 7
2 aid 8 7	2 aik 8 7	2 aim 8 7	2 aink 8 7
2 aitok 8 7	2 ak 8 7	2 am 8 7	2 atok 8 7
2 e 8 7	2 ed 8 7	2 ei 8 7	2 eid 8 7
2 eik 8 7	2 eim 8 7	2 eink 8 7	2 eitek 8 7
2 ek 8 7	2 em 8 7	2 etek 8 7	2 od 0
2 ok 0	2 om 0	2 otok 0	2 öd 8 7
2 ök 8 7	2 öm 8 7	2 ötök 8 7	2 uk 8 7
2 unk 8 7	2 ük 8 7	2 ünk 8 7	4 va 6
4 vad 6	4 vai 6	4 vaid 6	4 vaik 6
4 vaim 6	4 vaink 6	4 vaitok 6	4 vak 6
4 vam 6	4 vatok 6	4 ve 6	4 ved 6
4 vei 6	4 veid 6	4 veik 6	4 veim 6
4 veink 6	4 veitek 6	4 vek 6	4 vem 6
4 vetek 6	4 vuk 6	4 vunk 6	4 vük 6
4 vünk 6	7*10 d 2	7*10 i 2	7*10 id 2
7*10 ik 2	7*10 im 2	7*10 ink 2	7*10 k 2
7*10 m 2	7*10 nk 2	7 tek 2	7 tok 2
7 tők 2			

10

Határozószó

aktuálisan	alakilag	alaposan
alapvetően	automatikusan	azonnal
azonosan	aztán	általában
általánosan	belül	benn
bizonyára	egészen	egyaránt
egyedül	egyelőre	egyenként
egyértelműen	egyszer	egyszerre
előbb	elől	előre
elsősorban	elvileg	eredetileg
felfelé	feltehetőleg	felül
fenn	fokozatosan	függetlenül
grammatikailag	gyakorlatilag	gyakran
hamar	hasonlóan	határozottan
hatékonyan	hátul	heurisztikusan
hiába	hirtelen	ide
idén	időnként	ismét
itt	jelenleg	jelentősen
jobban	jól	jólláthatóan
keletebbre	később	kinn
kint	kizárólag	kizárólagosan
kívül	komolyan	korábban
korán	könnyedén	könnyen
közben	közvetlenül	különbözőképpen
különféleképpen	különösen	lassan
látszólag	legfőképpen	legközelebb
legmesszebb	legrészletesebben	legtöbbször
legvégén	lehetőleg	lényegesen
ma	magyarul	majd
maximálisan	már	máris
máshol	másképp	máskor
megint	meglehetősen	messze
még	mégis	mindig
mindjárt	míg	most
nagyrészt	nehezen	nemcsak
nevezetesen	néha	nyelvtanilag
nyilván	nyilvánvalóan	nyugatabbra
nyugodtan	olcsóbban	ott
pillanatnyilag	pontosan	potenciálisan
pozícionálisan	rendkívül	rendszerint
rég	régóta	részletesebben
rögtön	sehol	soha
sohasem	soká	szívesen
szükségképpen	szükségszerűen	tartalmilag
tavaly	tárgyasan	tárgyatlanul
távlatilag	tegnap	teljesen
természetesen	tényleg	többé
többféleképpen	többnyire	tradicionálisan
tulajdonképpen	túl	úgy



úgyszólván	újra	valahogy
valahol	valószínűleg	végleg
végre	végül	világosan

153 db Határozószó

11 Határozói igenév

belevéve	tévedvén	véve
----------	----------	------

3 db Határozói igenév

12 Határozói névmás

abba	abban	addig
ahhoz	ahogy	ahol
akkor	amerre	amint
amióta	amíg	amúgy
anélkül	annak	annál
arra	arról	attól
avval	azóta	azután
bárhogy	belé	ebbe
ebben	ebből	eddig
ehhez	ekkor	elé
emiatt	ennek	ennél
erre	erről	ezáltal
ezenkívül	ezután	felé
főlé	hogy	hogyan
hol	honnan	hova
hozzá	innen	így
közé	mellé	mielőtt
mindenképpen	miután	mögé
neki	oda	onnan
rá	semmiképpen	számára
ugyanahhoz	ugyanakkor	ugyanannak

63 db Határozói névmás

20 Igekötő

át	be	bele
együtt	el	ellen
elő	fel	fenn
hátra	hozzá	keresztül
külön	le	létre
meg	oda	össze
rá	tovább	túl
viszsa		

22 db Igekötő

38 Határozott névelő

a az  
2 db Határozott névelő

39 Határozatlan névelő

egy  
1 db Határozatlan névelő

40 Kötőszó

akár	avagy	azaz
azonban	ám	bár
de	és	ha
hanem	hiszen	hogy
illetve	is	különben
legalábbis	mert	mégsem
mind	mint	mintegy
mintha	mivel	míg
noha	pedig	például
pl	sem	sőt
tehát	továbbá	vagy
vagyis	viszont	

35 db Kötőszó

51 Igei módosító

hadd	hagy	ne
volna		

4 db Igei módosító

52 Melléknévi - igei módosító

abszolút	alig	annyira
egyre	elég	eléggé
éppen	főleg	igazán
igen	kevésbé	kissé
közismerten	majdnem	mind
nagyon	nem	nemigen
sem	szinte	talán
vajmi	viszonylag	

23 db Melléknévi - igei módosító

53 Számnévi módosító

kb	legalább	legfeljebb
mind	mindössze	mintegy

6 db Számnévi módosító

54 Névutószőrű szó

ellenére	keresztül	következtében
közben	közel	múlva
nyomán	óta	során
számára	szemben	túl

12 db Névutószőrű szó

60 Igeszerű szó

nincs	sincs
-------	-------

2 db Igeszerű szó

61 Nemigei segédige

érdemes	képes	köteles
muszáj	szabad	tanácsos
tilos		

7 db Nemigei segédige

62 Ragozhatatlan vagy ragozott főnévi névmás

engem	engemet	én
mi	ön	ő
ők	őt	sokan
te	téged	tégedet
ti		

13 db Ragozhatatlan vagy ragozott főnévi névmás

70 Egyéb szó

bizony	csak	csupán
egyáltalán	egyébként	esetleg
hát	igen	inkább
leginkább	mindazonáltal	mindegy
netán	persze	sajnos
stb	szintén	ugyan
ugye	vajon	

20 db Egyéb szó

100

Ige

abbahagy	ad	ajánl
akad	akar	alakít
alakul	alapít	alapsz
alapul	alkalmaz	alkot
alsz	analizál	azonosít
ábrázol	áll	állapít
állít	árt	átalakít
átalakul	átfog	bead
beépít	beérkez	beír
beleépít	beleért	belevesz
beprogramoz	beszél	bevezet
biztosít	bont	bonyolít
buk	céloz	csap
cserél	csinál	csoportosít
csökken	csökkent	definiál
derül	dolgoz	dönt
döntet	duzzad	egyesít
egyeztet	egyszerűsít	elalud
eldönt	elemez	elemz
eléged	elér	elhelyez
elindul	elkezd	elképzel
elkészít	elkészül	elkülönül
ellát	ellenőriz	elmúl
előállít	előáll	előfordul
előkészít	elront	elszakad
eltekint	eltér	elutaz
elvár	elveszít	elveszt
elvégez	emleget	említ
esz	él	épít
épül	ér	érez
érkez	ért	értelmez
érvényesül	fejez	feksz
felbont	feldolgoz	felépít
felépül	felhasznál	felismer
felmerül	felold	felsorol
felszerel	feltár	feltételez
feltölt	felül	felvetőd
fennáll	fél	fér
figyel	fog	fogalmaz
foglal	foglalkoz	folly
folytat	fordít	fordul
függ	fűz	gátol
generál	gondol	gondoskod
gyűjt	hagy	hajt
halad	hall	hallgat
használ	használtat	hat
határol	határoz	hazaugr
hazautaz	hátravet	hiányz
hisz	hív	hoz
hozzácsatol	hozzárendel	húz
igazít	iktat	illet

implementál	indul	irányul
ismer	ismertet	ir
javít	jár	jelen
jelent	jelez	jelöl
jelz	jön	jut
kap	kapcsol	kapcsolód
keres	kerül	kezel
képez	képz	képzél
kér	kérdez	kés
késleltet	készít	készítet
készül	kiad	kialakít
kialakul	kiderül	kidolgoz
kiegészít	kielemez	kielemez
kiérkez	kigyűjt	kihagy
kihelyez	kiindul	kiismer
kiír	kijavít	kiküld
kiküszöböl	kiokoskod	kirajzolód
kisajátít	kisérel	kisérl
kitér	kitisztít	kiutaz
kiválaszt	kizár	kíván
koncentrál	korlátoz	kódol
köt	kötőd	következ
küld	különböz	különböztet
lassít	lát	látsz
leáll	lefordít	legy
legyen	leír	leledz
lelombozód	len	lesz
leválaszt	lép	létez
magábarejt	magyaráz	marad
megad	megállapít	megcáfol
megcéloz	megcsappan	megegyez
megelégsz	megelőz	megemlít
megérkez	megért	megfelel
meghalad	meghatároz	megindít
megismerked	megismétel	megismétl
megjelen	megjelöl	megjósol
megkap	megkeres	megkérdőjelez
megkülönböztet	megnevez	megnyom
megold	megoszl	megőriz
megőrz	megpróbál	megszok
megszűn	megtalál	megtanul
megten	megtesz	megvalósít
megvan	megvizsgál	megy
men	mond	módosít
mutat	mutatkoz	működ
működtet	nehezít	nevez
nő	nyilvánul	nyújt
odaér	okoz	old
olvas	oszt	összead
összeállít	összefoly	összegez
összegyűjt	összegz	összekapcsol
összeköt	összevon	produkál
programoz	próbál	ragaszt
ragoz	rajzol	rak
rakód	rájön	rájöv



reagál	redukál	rendel
rendelkez	reprezentál	rögzít
segít	sorol	szab
szakad	számít	számol
szán	származ	szemantizál
szemléltet	szenved	szerepel
szeret	szerez	szervez
szerz	szolgál	szorít
szól	szület	talál
találkozik	tanul	tanulmányoz
tart	tartalmaz	tartoz
támaszkodik	támaszt	támogat
tán	tárgyal	tárol
te	tekint	ten
teremt	tervez	tesz
tevőd	tételez	toldalékol
torzít	továbbfejleszt	továbbvin
továbbvisz	tökéletesít	töreked
törekész	törőd	történ
tud	tükröz	tűn
tűz	un	utasít
utaz	vagy	valósít
van	vál	válaszol
választ	változ	vár
ven	vesz	veszít
vet	vetőd	vezet
vezérel	vezérl	végez
végignézi	végigolvas	végigpróbál
végrehajt	végz	vél
világít	vin	visszafordít
visszajön	visszakérdez	visszaküld
visszaszól	vizsgál	volt
vonatkozik	zavar	
416 db Ige		

110 Segédige

akar	fog	kezd
kíván	próbál	szeret
szok	talál	tud

9 db Segédige

120 Féligéi segédige

ill	kell	lehet
sikerül		
4 db Féligéi segédige		

201 Név

Apreszján	április	Bach
Béla	Budapest	Chomsky
COSY	Dezső	DOS
Dubna	Eötvös	Erna
Ernő	Eszenszki	február
György	János	József
Kantor	KFKI	KGST
Kovács	Kurcsatov	Lóránd
május	március	Mátyás
Moszkva	Naszódi	Páll
Péter	Pista	
32 db Név		

202 Mértékegység

fő	Mbyte	méter
tucat		
4 db Mértékegység		

210 Főnév

adat	adatbázis	adatkapcsolat
adatkészlet	akció	alak
alany	alanyeset	alap
alapforma	alapszó	aláírás
alfejezet	algoritmus	alkalmazhatóság
alkalm	alkalom	alkotórész
alrendszer	alternatíva	analizálás
analízis	anyacég	anyag
anyagnév	argumentum	aspektus
attribútum	azonosítás	ábra
ábrázolás	ágens	ágy
állapot	álláspont	állítmány
általánosság	átalakítás	átalakulás
átállítás	átmenet	áttekintés
átütemezés	átvétel	baj
beépítés	beiktatás	bejelentkezés
betű	bevezetés	bevétel
bél	bika	birtok
birtokjel	birtoklás	birtokosjel
birtokviszony	bizonyítás	bizottság
bővítmény	Bulgária	cél
célgép	célkitűzés	címszó
csap	Csehszlovákia	cselekvés
csomó	csoport	csoportvezető
csökkenés	csőd	csütörtök
darab	definíció	dialógus
diszk	dolg	dolog
egyesszám	egyetemista	egyezmény

egyeztetés	egység	egységesítés
elem	elemzés	elírás
eljárás	elképzelés	elkészítés
ellenőrzés	elmélet	előállítás
előírás	előny	előtag
eltérés	elv	elválsztás
ember	eredmény	eredményül
esemény	eset	esetenként
esetgrammatika	esetrag	esetragozás
esz	eszköz	eszközül
élőlény	ér	érdek
érdekeltség	érdekesség	érdeklődés
érdem	értelem	értelm
értelmezés	értelmezhetőség	érték
érthetőség	ész	ételrecept
év	évtized	fajta
fázis	fejezet	fejletlenség
fejléc	feladat	felbontás
feldolgozás	felépítés	felfogás
felismerés	feloldás	felsorolás
felsőfok	felszóítómód	feltevés
feltételesmód	féleség	figyelem
figyelembevétel	figyelm	fizika
fizikus	fogalm	fogalom
fogás	fok	fokozás
folyam	fordítás	forma
formula	forrás	fókusz
földarabolás	főnev	főnév
főnévrag	frázis	függőség
függvény	generáció	generálás
generátor	gép	gépírás
gomb	gond	grammatika
gráf	gyakoriság	gyakorlat
gyengeség	gyerek	gyermek
gyermekbetegség	gyorsaság	gyökér
gyökér	halmaz	hang
hangsúlyozás	használat	határozó
hatás	hatékonyság	hálózat
hátrány	háttér	háttértároló
hely	helyesség	helyhatározó
helyzet	hét	hiány
hiánycikk	hiba	hibajelzés
hipotézis	hivatkozás	hívás
homlokter	homloktér	homonímia
hozzárendelés	hó	idő
időhatározó	ige	igealak
igeidő	igekötő	igenev
igenév	igerag	igeragozás
igénként	igény	illeszkedés
indikáció	indoklás	infinitivusz
információ	információközlés	input
intelligencia	interfész	interfészül
interrupt	intézet	intézmény
irány	irányvonal	irányzat
irodalm	irodalom	iskola

ismeret	írásjel	javítás
járás	jegy	jel
jelenidő	jelentés	jelentőség
jelleg	jelölés	jelzés
jelző	jogász	jogosultság
jogtudomány	jövő	jövőidő
kabát	kalandozás	kalap
kapcsán	kapcsolat	kapcsolódás
karakter	kategória	katona
kedv	kelet	kereset
keresés	keret	kezdet
kezelés	kép	képesség
képlet	kérdés	kés
készítés	készlet	kétértelműség
kéz	kiadás	kialakítás
kicserélés	kidolgozás	kiegészítés
kifejezés	kikeresés	kiküldetés
kinyerés	kiutazás	kiút
kiválasztás	kivétel	Kína
kíséret	kísérlet	kombináció
komunikáció	korlát	korlátozás
kód	kölcsönhatás	költemény
könyv	körülmény	kötél
kötél	kötőhang	kötőjel
követelmény	közép	közkincs
központ	köz	Kuba
kutatás	különbség	különlegesség
kvantor	labor	laboratórium
látóhatár	láz	lefordítás
legelő	lehetőség	leírás
lekérdezés	lekszéma	Lengyelország
leválasztás	level	levél
lé	lényeg	lépés
létrehozás	lista	logika
ló	magánhangzó	Magyarország
manipuláció	maradék	matematikus
mássalhangzó	megadás	megállapítás
megértés	meghatározás	megindoklás
megjegyzés	megkeresés	megközelítés
meglét	megnevezés	megnyomás
megoldás	megtalálás	megvalósítás
megye	mellékmondat	melléknev
melléknév	memória	menet
mennyiség	mennyiségjelző	mentesség
menü	mező	méret
mérőrendszer	mérték	mértékegység
milyenség	minta	mondanivaló
mondat	mondatrész	mondattan
Mongólia	morféma	morfológia
mód	módszer	mult
multidő	munka	munkahipotézis
munkatárs	munkás	mutatónévmás
működés	működőképesség	műnyelv
nagyságrend	nehézség	nev
név	névelő	névmás

névszó	névutó	nézet
növelés	nő	nyelv
nyelvész	nyelvész	nyelvmodell
nyelvtan	nyelvtudomány	nyugat
objektum	ok	oldal
olvasás	ország	osztály
önállóság	összeállítás	összeg
összekötés	ősz	papír
paraméter	parancs	park
párbeszéd	példa	péntek
piac	pillanat	pont
pontpár	pozíció	pozitívum
preferencia	prepozíció	probléma
processzor	profil	program
programozás	programrendszer	programrész
projekt	prompt	rag
ragozás	reagálás	remény
rendelkezés	rendszer	rendszerterv
rész	részeredmény	részhalmoz
részlet	részletesség	résznyelv
részvényes	rétég	robbanás
Románia	rovat	rugalmasság
ruha	sablon	sajátosság
sebesség	segédige	segédzó
segítség	siker	sikertelenség
sokaság	sor	sorozat
sorrend	specifikáció	státusz
súly	szabály	szakasz
szakember	szakértő	szakirodalom
szakképzettség	szarv	szál
szám	számítástechnika	számítástechnikus
számítástudomány	számítógép	számjel
számjelző	számnev	számnév
számológép	szándék	szelekció
szellem	szemantika	személy
személynév	személyrag	szeminárium
szempont	szerep	szerkezet
szervíz	szerződés	szétvágás
szimbólum	szinoníma	szint
szintagma	szintaxis	szituáció
szokás	szombat	Szovjetunió
szó	szócska	szócsoporth
szófaj	szókészlet	szóosztály
szórend	szótár	szótő
szöveg	szövegrész	szövetkezet
szükség	születés	találkozás
találkozó	tanulmány	tanulmányozás
tapasztalat	tartalm	tartalmiság
tartalom	tábla	táblanev
táblanev	táblázat	tájékoztató
támogatás	tár	tárgy
tárgyasság	tárgyeset	tárgyrag
tárolás	távolság	technika
teherterhel	tekintet	termék
terminológia	terület	terv



tevékenység	téesz	téma
témakör	tér	tisztázás
típus	toldalék	tó
többszám	többsértelműség	többség
tőtár	tudat	tudás
tudományegyetem	tudomás	tulajdonság
tulajdonságjelző	tündérmese	tűz
univerzalitás	utalás	utasítás
utca	utibeszámoló	utód
újdonság	újraalkalmazhatóság	út
üzenet	valószínűség	vasárnap
válasz	válaszidő	választék
vállalat	változat	változó
vásár	vektor	veszély
vesztés	vezérlés	vég
végeredmény	végignézés	végigpróbálgatás
végpont	végrehajtás	végződés
vélemény	vér	Vietnam
visszahatás	visszakeresés	vita
vizsgálat	vonal	vonás
vonzerő	zaj	zavar

642 db Főnév

211 Egyszerre főnév és melléknév

alkotó	angol	automata
beszámoló	beteg	birtokos
bulgár	default	dolgozó
egyenes	fa	felhasználó
folyó	francia	illető
japán	kalcium	kálium
képző	konzerv	lengyel
magyar	modell	nyugatnémet
olasz	orosz	programozó
szövegszerkesztő		

28 db Egyszerre főnév és melléknév

212 Főnév jellegű névmás

aki	ami	az
efféle	ez	ilyesmi
ki	másik	mi
mindaz	mindenki	mindez
semmi	senki	többi
ugyanaz	ugyanez	valaki
valami		

19 db Főnév jellegű névmás

213 Személyragozott névmás

egyik	jómag	mag
önmag		
4 db Személyragozott névmás		

214 Birtokosan nem ragozódó névmás

egymás	enyém	mind
3 db Birtokosan nem ragozódó névmás		

220 Melléknév

absztrakt	absztraktabb	agglutinatív
ajánlatos	akadémiai	aktuális
alacsony	alacsonyabb	alakú
alapú	alapvető	alábbi
algoritmikus	alkalmas	alkalmi
aprólékos	aspektusú	atomkutató
automatikus	automatizálási	azonos
állandó	általános	általánosabb
átlagos	belső	bemeneti
bizonyos	biztos	célratörő
célszerű	divatos	divatosabb
dubnai	eddigi	egész
egyedi	egyenértékű	egyértelmű
egységes	egyszerű	egyszerűbb
ellenkező	elméleti	előbbi
előnyös	előtti	előző
első	elvi	eredeti
erős	európai	explicit
érdekes	érdektelen	érdemi
értelmes	értelmesebb	értelmű
érvényes	ésszerű	felesleges
felszíni	feltételes	fenti
finom	finomabb	fogalmi
fontos	formális	fő
főnévi	fős	furcsa
független	függőségi	gazdag
gazdaságos	generációs	geometriai
gépi	gépikódú	gondos
grafikus	grammatikai	gyakori
gyakorlati	gyenge	gyors
halvány	hangtani	hasonló
hasonszerű	használatos	hasznos
haszontalan	határozatlan	hatékony
hatékonyabb	helyes	helytelen
heurisztikus	hiányos	hibás
hivatott	homályos	hosszabb
hosszú	ideális	idegen
igaz	igei	igényű
implicit	indirekt	infinitivuszi

interaktív	ismeretlen	jelen
jelenidejű	jelentésű	jelentős
jellegű	jelzői	jelzős
jobb	jó	jólstruktúrált
kapcsolatos	karakteres	kategóriájú
káros	kellemetlen	kereseti
kezdeti	kék	képzettségű
kérdéses	kész	kétértelmű
kétféle	kicsi	kifejezési
kijelentőmódú	kintdolgozó	kinti
kis	kisebb	kiváló
kívánatos	konkrét	korábbi
korrekt	könnyebb	könnyű
kötetlenebb	közbülső	közismert
köznapi	közös	közvetlen
külföldi	külső	laza
legegyszerűbb	legfőbb	lehetetlen
lehetőséges	lexikai	lexikális
lényeges	lényegi	logikai
magas	magasabb	matematikai
maximális	márciusi	más
meglehető	megszámlálhatatlan	melléknévi
mesterséges	mély	méretű
mindennapi	minimális	moduláris
mondattani	morfológiai	multidejű
nagy	nagyobb	nagyobbméretű
nehéz	nehéz	nemzetközi
nevű	német	névelőtlen
névelős	névszói	normális
nyelvészeti	nyelvi	nyelvtani
nyelvű	nyilvánvaló	objektív
opcionális	osztatlan	önálló
párbeszédes	piros	pontatlan
pontatlanabb	pontos	pontosabb
pontszerű	portábilis	potenciális
pozicionális	praktikus	profi
programfüggetlen	programozástechnikai	ragos
ragtalan	reális	redundáns
releváns	rettenetes	részleges
részletes	rokonszenves	rossz
rózsaszín	rövid	rövidtávú
sajátos	sajnálatos	sikeres
sokféle	sorrendi	speciális
specifikációs	statikus	stílusú
strukturális	súlyos	szabad
szaggatott	szakmai	számítástechnikai
számítógépes	számú	százszoros
szelektív	szemantikai	szemantikus
személyes	személyű	szerű
szigorú	szimmetrikus	szintaktikai
szintaktikus	szintű	szocialista
szokásos	szovjet	szórendű
szöveges	szükséges	szükségszerű
születési	tajvani	tartós
tájékoztalan	tárgyas	tárgyatlan

tárgyragos	tárgyú	távoli
tele	teljes	tematikus
természetes	természetű	terminológiai
tetszőleges	téves	típusú
további	többes	többesszámú
többértelmű	tradicionális	tudományos
túlsó	un	univerzális
utolsó	utóbbi	úgynevezett
új	újabb	valódi
valószerű	változatlan	váratlan
végtelen	világos	vonalszerű
zavaros		

316 db Melléknév

221 Melléknévi igenév

adható	alakítható	alkalmazandó
alkalmazható	azonosítható	állítható
álló	bedolgozó	beszélő
bevezethető	bővíthető	cselekvő
cserélhető	el döntendő	elemző
elérhető	elfogadható	elképezhető
elkészítendő	ellenőrizhető	ellenőrzendő
elváló	elvégezhető	elvégzendő
élenjáró	érhető	értetődő
fekvő	felbontható	felhasználható
felszólító	feltételezhető	fentemlített
félő	figyelhető	fogható
foglалható	foglalkozó	fordító
függő	gyorsítható	gyorsító
hallgató	használandó	használható
határozó	hiányzó	hozzáilleszthető
hozzátartozó	idevonatkozó	idéző
illeszthető	javítható	járó
jellemző	jelző	jósolható
jövő	kapcsolható	kapcsolódó
kapható	kedvező	kezelő
képezhető	kérdő	készítendő
kiadandó	kidolgozandó	kielégítő
kiérkezendő	kigyűjtő	kiinduló
kijelentő	kiolvasható	kiváló
kötelező	következő	kutató
különböző	látható	látászó
leíró	lekérdező	levő
létező	megelőző	megfelelő
megjelenő	megjósolható	megoldható
megtalálható	megtehető	menő
módosító	működtethető	működtető
műveltető	nyelvelemző	nyugtalanító
olvasható	összetartozó	összetett
összetevő	résztevő	rögzíthető
segítő	sorolható	szenvedő
<b>szerelő</b>	szolgáló	szóló
található	találkozó	tartozó

230 Szám

egy	egyetlen	egyszer
először	első	fél
félíg	harmadik	hat
három	hét	kettő
két	második	nulla
öt	ötödik	

17 db Szám

231 Melléknév jellegű számnév

csekély	elegendő	kevés
sok	számos	több
többször		

7 db Melléknév jellegű számnév

232 Számnévi névmás

ahány	akárhány	akármennyi
amennyi	annyi	bármennyi
csomó	ennyi	mennyi
mindkét	néhány	némi
pár	sehány	semennyi
számtalan	valahány	valahányszor
valamennyi		

19 db Számnévi névmás

233 Kvantor

akármelyik	amely	amelyik
bármely	bármelyik	egyik
mely	melyik	mindegyik
minden	összes	semelyik
valamelyik		

13 db Kvantor



tájékoztató	tároló	tekinthető
továbbosztható	történő	tudható
túlmenő	utaló	való
valósítható	választható	velejáró
vezérlő	végrehajtó	viSSZautaló
vonatkozó		
130 db Melléknévi igenév		

222 Melléknévi névmás

akkora	akkori	amolyan
bármilyen	egyes	egyéb
ily	ilyen	milyen
mindenféle	olyan	saját
semmilyen	ugyanolyan	valamilyen

15 db Melléknévi névmás

223 Melléknévi névutó

beli	feletti	közti
nélküli	szerinti	
5 db Melléknévi névutó		

alatt	alá	alól
által	belé	belől
belül	benn	elé
ellen	elől	előtt
ért	felett	felé
felől	felül	fölé
fölött	helyett	hozzá
kívül	közé	között
közt	közül	mellett
mellé	mellől	miatt
mögé	mögött	mögül
nál	nek	nélkül
rajt	ra	ról
szerint	től	után
vel		

